



Analysis of Colorectal Cancer Screening Regimens

ROBERT T. CLEMEN *

Fuqua School of Business, Duke University, Box 90120, Durham, NC 27708-0120, USA
E-mail: clemen@mail.duke.edu

CHRISTOPHER J. LACKE

Mathematics Department, Rowan University, Glassboro, NJ 08028, USA
E-mail: lacke@rowan.edu

Received 16 August 2000; Revised 9 March 2001

Abstract. We analyze several colorectal cancer screening methods. We begin with an existing deterministic model of the colorectal cancer growth-and-development model. Using judgments from two knowledgeable experts on colorectal cancer, we incorporate probability distributions for important parameters in the model. The analysis proceeds in three phases: First is a straightforward Monte Carlo simulation that includes uncertainty about structural parameters, the results of which identify five dominant screening strategies in terms of the expected number of cancers prevented and expected cost per life-year saved. The next part of the analysis develops a two-attribute utility function to rank order the screening regimens. The results show the same top five, with the top-ranked strategy being colonoscopy every three years. Sensitivity analysis demonstrates the robustness of the results.

1. Introduction

Colorectal cancer (CRC) is one of the four most prevalent forms of cancer in the United States. According to the American Cancer Society [1], physicians will diagnose over 135,400 new cases of invasive colorectal cancer in 2001. The American Cancer Society further predicts that CRC will cause over 56,700 deaths during 2001. CRC responds well to treatment, though; 61% of treated patients survive more than five years, as compared to 14% for untreated patients. Patients treated for localized CRCs have a survival rate of almost 90% [2].

Good screening procedures contribute to CRC's high survival rates. Moreover, these procedures detect benign adenomatous polyps, the primary precursors to CRC. Surgeons can remove such polyps before they become cancerous lesions. Screening methods include four procedures. Three are invasive: flexible sigmoidoscopy (FSIG), colonoscopy (CSCPE), and double-barium contrast enema (DCBE). The fourth is the non-invasive fecal occult blood test (FOBT).

Choosing a screening regimen is an important health-policy problem, but the choice requires difficult trade-offs. Test costs range from five dollars for FOBT to approximately \$1000 for CSCPE. In addition, a small percentage of invasive procedures result in bowel perforation, requiring surgical repair and occasionally resulting in death. In some ways, though, invasive procedures are preferable to FOBT. Because FOBT can detect blood in the stool, which can result from many sources, FOBT has a high probability of a false positive. A positive result requires a diagnostic colonoscopy, and so a false positive can be costly in both

financial and emotional terms. In addition, FOBT will not detect non-bleeding polyps and may fail to detect a bleeding lesion. Hence FOBT has large error rates. In contrast, invasive procedures are more accurate because they view at least part of the colorectal region.

Cancer researchers create models of the CRC development-and-growth process and use them to evaluate alternative screening regimens. In this paper, we focus on such a model described by Wagner et al. [3]. Wagner's model is a cost-effectiveness model that estimates a variety of outputs, including the number of colorectal cancers and late stage cancers prevented by early detection, the cost of a screening regimen, and the cost per life year saved. Inputs include polyp and cancer incidence rates, test sensitivity and specificity, mortality rates, stage-specific dwelling times and procedural costs. The model calculates the outputs for 100,000 asymptomatic, average risk individuals, aged 50–85, in the screening program, assuming complete compliance with surveillance and follow-up programs.

Like many such models in the literature, Wagner's is deterministic. That is, it ignores many uncertainties associated with the cancer-development process. In particular, even though input parameters are estimated on the basis of data or judgment, Wagner's analysis of the model assumes that the parameters are fixed and known.

We enhance Wagner's model by incorporating uncertainty about eight of the input parameters. Because data for estimating the input parameters are sparse, we rely on judgments from two experts to provide the necessary input probability distributions. For each parameter, two experts separately provided subjectively assessed distributions, which we combine using Jouini and Clemen's copula-based

* Corresponding author.

method [4], thereby creating the input distributions needed for a full probabilistic risk analysis of the various screening regimens. With our modifications, the model becomes one in which each individual in the subject population is viewed as having the same parameters (e.g., the risk of cancer at age 55), but the parameter value is unknown.

Our risk analysis proceeds in three phases. First is a straightforward Monte Carlo simulation, which identifies five screening regimens that dominate others in terms of the mean number of CRCs prevented and the mean cost per life-year saved (CPLYS). These five include DCBE every three years; CSCPE alone every three or five years; and CSCPE every three or five years coupled with an annual FOBT. This initial analysis, however, cannot produce a full rank ordering of regimens; for that, an analysis that incorporates a trade-off between CRCs prevented and CPLYS is required. Phase 2 of the analysis develops a utility function for these two attributes and uses it to produce a ranking of the screening regimens. This analysis identifies CSCPE alone every three years as the top choice, followed by CSCPE every three years with an annual FOBT. Finally, we perform a sensitivity analyses on the trade-off weight used in the utility function, the assumed sensitivity of DCBE, and the correlation used in the aggregation procedure.

The next section describes Wagner's model and the twenty-five screening regimens that we study. Sections 3 and 4 present the probabilistic model and our analysis results, respectively. Section 5 concludes with a discussion of the approach and results as well as suggestions for further research.

2. A model of colorectal cancer

Our work begins with Wagner et al.'s [3] cost-effectiveness model for CRC screening regimens. Wagner's model assumes that people enter a screening program at age 50, receiving a regimen of screening processes at some regular interval or combination of intervals.

Each regimen that includes an FOBT uses the test on an annual basis. Invasive procedures are used at intervals of 3, 5, or 10 years, with participants receiving the invasive procedure at age 50 unless the screening regimen includes FOBT along with the invasive test. A positive FOBT, FSIG, or DCBE is assumed to be followed by a diagnostic colonoscopy. A polyp found during colonoscopy (screening or otherwise) can be removed immediately. Individuals having polypectomy or treatment for CRC then enter a surveillance program using colonoscopy. After age 85, a person leaves the screening program; however, surveillance continues for those in the surveillance program. Cessation of screening at age 85 reflects the notion that a person who has had no polyps or CRC by age 85 is more likely to pass away from other causes before a polyp could develop and become a late stage CRC.

In order to standardize results, all calculations are done in terms of a group of 100,000 50-year-old, average-risk people. In each year, each person can be in one of five

states of health: (1) alive without polyps or CRC; (2) alive with adenomatous polyps; (3) alive with early stage CRC; (4) alive with late stage CRC; and (5) dead. Each time a test is performed, various outcomes are possible. A negative result, whether correct or not, results in the person remaining a member of the screening group. A false positive also returns the individual to the screening group, following the diagnostic colonoscopy associated with the positive result. A true positive places the individual in the surveillance group. However, each time an invasive procedure is performed, the patient's colon may be perforated, requiring surgical repair. Unfortunately, a percentage of these patients do not survive the procedure. Also, some patients who undergo colonic resection to treat CRC do not survive, so the people enter and exit the surveillance group each year. Another group of people have "lifetime-latent" cancers, CRC's that progress so slowly that they would not be diagnosed clinically in the absence of screening. Due to the screening process, though, these cancers are detected and treated, which means that these individuals are treated and enter surveillance unnecessarily.

An individual's progression from one state to the next is based upon expected value calculations. For example, suppose the sampled value for the prevalence of polyps at age 50 is 0.2. Then 80,000 (out of the original 100,000) people are expected to be free of polyps and 20,000 are expected to have polyps. Suppose that a sigmoidoscopy will be used in the first year. Then, using an effective reach of 50% of the colon, 10,000 people are expected to have polyps in the region of the colon that can be viewed by the sigmoidoscope. Using the values for sensitivity and specificity, the expected numbers of positives and negatives (true and false) are calculated, after which the expected treatment costs and life years saved are determined. Those people who do not enter the surveillance program and do not die comprise the group at age 51. At this point similar calculations are made, keeping track of the people with polyps that have gone undiscovered and using annual polyp incidence rates to determine how many of the people who were clear of polyps at age 50 develop them at age 51. This procedure continues through age 85.

The twenty-five screening strategies that we evaluate include all of those covered in [3], plus a seven-year interval for the invasive procedures, thereby permitting screening of individuals when they exit the program at age 85. Thus, each screening strategy we considered includes one of the three invasive procedures at 3, 5, 7, or 10 years, with or without an annual FOBT, or solely using an annual FOBT with no invasive procedure.

We made some minor modifications to Wagner's model. Survival rates based on 1973–1993 data from the Surveillance, Epidemiology, and End Results (SEER) project were included to update the model. As suggested by Lipscomb [5], the discount rate for future years of life was reduced from 5 to 3%. (Sensitivity analyses performed on the discount rate (5 versus 3%) showed no change in the final results.) Because the model was modified under the auspices

of the North Carolina Central Cancer Registry, we tried to make the model as specific to North Carolina as possible. In particular, incidence rates for CRCs were obtained from the North Carolina Central Cancer registry; age-specific non-CRC mortality rates for 1990–1995 came from the North Carolina Center for Health Statistics; and unit costs for testing came from cost information collected by Medicode [6] for specific zip codes within North Carolina. For a complete description of the modified model, see [7].

3. An assessment of structural uncertainty

3.1. Variable selection

Based on the results of sensitivity analyses reported by Wagner et al., as well as discussions with other colorectal cancer experts, the following eight parameters were chosen for probabilistic modeling:

- (1) Sensitivity of FOBT, that is, the proportion of positives that are correctly identified.
- (2) Proportion of the population with adenomatous polyps at age 50.
- (3) Proportion of CRCs that develop from adenomatous polyps.
- (4) Proportion of early stage CRCs that would be discovered in the absence of any screening (i.e., the proportion that would become symptomatic while in the early stage).
- (5) Expected dwelling time (months) for early stage CRC (i.e., the length of time that it remains early stage).
- (6) Expected dwelling time (months) for late stage CRC.
- (7) Expected dwelling time (months) for an invasive procedure (i.e., the amount of time from when a benign adenomatous polyp is just large enough to be seen using sigmoidoscopy until it becomes a malignant tumor).
- (8) Expected dwelling time (months) for FOBT (i.e., the amount of time from initial bleeding until an adenomatous polyp becomes a malignant tumor).

Although they represent proportions (variables 1–4) and expected dwelling times (variables 5–8), all eight of these variables are treated in the model as uncertain parameters. Thus, our probabilistic model is one in which we view all members of the population as having the same parameter, but those parameters are unknown. Each iteration in the Monte Carlo simulation randomly chooses a value for each of the eight uncertain parameters and makes the necessary calculations with those values.

3.2. Expert probability assessment

The experts selected to provide probability distributions for the eight parameters were Dr. Dawn Provenzale of the Duke

University Medical Center and Dr. Robert Sandler of the Sheps Center at the University of North Carolina at Chapel Hill. Each expert was interviewed separately to obtain subjective probability judgments for the eight parameters. Each interview lasted approximately 2.5 h and followed conventional probability-elicitation procedures [8–10]. Using standard techniques, the median, first and third quartiles, minimum and maximum, and 0.05 and 0.95 fractiles were obtained for each variable. Table 1 shows elicited cumulative probabilities for the eight variables.

Examination of table 1 reveals that, in very general terms, the experts are in agreement regarding most of the variables, especially the sensitivity of FOBT (variable 1) and the proportion of CRCs that develop from adenomatous polyps (variable 3). For several of the variables, though, the two experts disagree regarding one or both of the tails. For example, consider the expected dwelling time for early-stage CRCs; expert 1 puts 90% of the probability mass between 9 and 34 months, whereas expert 2 puts 90% of the mass between 20 and 46 months.

3.3. Fitting distributions

For modeling convenience we fit parameterized distributions to the assessed distributions in table 1. Before going through the fitting procedure, though, we modified the distributions slightly so that each pair would have the same support interval. (We did this in order to ensure reasonable results from the aggregation procedure; details below.) For variables 1–4 (the proportions), the assessed upper and lower bounds were

Table 1
Cumulative probabilities assessed by two experts. For confidentiality reasons, the two experts are not identified with their specific responses.

Variable	Fractile						
	0.00	0.05	0.25	0.50	0.75	0.95	1.00
1. Sensitivity of FOBT							
Expert 1	0.01	0.03	0.08	0.13	0.16	0.19	0.20
Expert 2	0.005	0.01	0.02	0.05	0.09	0.13	0.15
2. Proportion with polyps at age 50							
Expert 1	0.05	0.10	0.33	0.40	0.47	0.56	0.60
Expert 2	0.20	0.25	0.29	0.35	0.40	0.55	0.60
3. Proportion of CRCs from polyps							
Expert 1	0.70	0.77	0.82	0.90	0.94	0.98	0.99
Expert 2	0.80	0.85	0.88	0.90	0.925	0.94	0.95
4. Proportion of discovered early-stage CRCs							
Expert 1	0.10	0.11	0.15	0.18	0.20	0.31	0.33
Expert 2	0.05	0.07	0.17	0.25	0.33	0.43	0.45
5. Expected dwelling time for early stage CRC (months)							
Expert 1	6	9	18	24	30	34	36
Expert 2	18	20	36	40	42	46	48
6. Expected dwelling time for late stage CRC (months)							
Expert 1	6	8	16	18	21	23	24
Expert 2	6	8	15	24	30	35	36
7. Expected dwelling time for invasive procedure (months)							
Expert 1	48	54	63	66	73	81	84
Expert 2	24	30	46	60	74	90	96
8. Expected dwelling time for FOBT (months)							
Expert 1	0	2	6	12	18	22	24
Expert 2	0	2	4	6	8	10	12

Table 2
Parametric models fit to the experts' elicited probabilities.

Variable	Expert 1	Expert 2
Sensitivity of FOBT	Logistic(0.12, 0.0383)	Gamma(1.48, 0.0426)
Proportion with polyps at age 50	Logistic(0.38, 0.0768)	Gamma(13, 0.028)
Proportion of CRCs from polyps	Beta(11.20, 1.55)	Weibull(24.32, 0.91)
Proportion of discovered early-stage CRCs	Lognormal(0.19, 0.0651)	Weibull(2.26, 0.28)
Expected dwelling time for early stage CRC	Weibull(3.34, 26.00)	Beta(2.82, 0.98)*42 + 6
Expected dwelling time for late stage CRC	Logistic(17.75, 2.80)	Beta(1.06, 0.86)*30 + 6
Expected dwelling time for invasive procedure	Logistic(66.975, 5.72)	Beta(1.37, 1.37)*72 + 24
Expected dwelling time for FOBT	Beta(1.11, 1.11)*24	Logistic(6.3, 1.94)

assumed to be the 0.01 and 0.99 fractiles, and we set the minimum and maximum for each of these variables to be zero and one. For each dwelling-time variable, the lower bound for each expert was set at the lower of the 0.00 fractiles from the two distributions, and likewise, the upper bound for each expert was set at the greater of the two 1.00 fractiles. Thus, for example, for the expected dwelling time for early-stage CRCs (variable 5), we set the lower bound for each expert at 6 months and the upper bound at 48 months.

With these adjustments to the assessed data, we used BestFit [11] to find best-fitting parametric models for each distribution, using the Kolmogorov–Smirnov statistic as the fitting criterion. The results are shown in table 2, with the density functions, means, and variances of the distributions outlined in appendix A. Graphs of these fitted distributions were shown to the experts and interpreted for them, and they agreed that these models appropriately represented their beliefs about the input parameters.

3.4. Combining distributions

A large literature exists on the aggregation of probability distributions. See [12–14] for critical reviews. The simplest solution is to simply average the two distributions. For example, if we have distributions $F_1(x)$ and $F_2(x)$, we can calculate the average $\bar{F}(x) = [F_1(x) + F_2(x)]/2$. Although such averaging is common practice, it ignores the relative accuracy of the experts as well as potential dependence between the experts. In the case of our experts, there was no reason to believe one was more accurate than another. Expert judgment can be thought of as probabilistically dependent in many different senses; for our purposes, we think of dependence in terms of errors in estimation or forecasting tasks. That is, positive dependence occurs when the experts tend to make errors of the same sign. (This concept is made precise in the context of the model described in appendix B.) Both experts in this study mentioned results from the same polyp study, attended the same medical school, and had consulted each other previously, leading us to believe that dependence was a real issue that should be accounted for in the aggregation procedure.

To account for the dependence, we used Jouini and Clemen's Bayesian copula-based approach [4] to aggregate the distributions. The model is Bayesian in the sense that it takes the perspective of a decision maker (DM, sometimes called a "supra Bayesian") who has input distributions from

experts and must use it to arrive at a posterior distribution. Formally, the DM's problem becomes one of constructing a likelihood function that relates the information from the expert. Whether the DM has an informed prior or not, ultimately the problem becomes one of applying Bayes' theorem to the likelihood function and prior in order to derive the posterior distribution. The reviews mentioned above [12–14] discuss the Bayesian approach in more detail.

In Jouini and Clemen's model, the likelihood function is constructed using a copula to encode the dependence among the experts. Jouini and Clemen give a complete and formal description of their approach; we describe it in appendix B less formally and, we hope, in a way that is more accessible. Jouini and Clemen's result is that the DM's posterior distribution, given expert distributions F_1 and F_2 (and corresponding densities f_1 and f_2), can be written as

$$f_{DM}(x|f_1, f_2) = f_1(x)f_2(x)c[1 - F_1(x), 1 - F_2(x)], \quad (B6)$$

where $c[\cdot]$ is a copula, a function that encodes the dependence between the expert assessments. For complete details and further references on copulas, consult appendix B.

3.5. Eight combined distributions for the colorectal cancer model

Implementing the copula-based aggregation model in (B6) is straightforward. As described in appendix B, we take c to be the copula density that underlies the multivariate normal density. Thus, the formula for combining $f_1(x)$ and $f_2(x)$ is

$$\begin{aligned} f_{DM}(x|f_1, f_2, r_X) &= f_1(x) \times f_2(x) \\ &\times \exp\{-r_X(r_X(\Phi^{-1}[1 - F_1(x)])^2 \\ &- 2\Phi^{-1}[1 - F_1(x)]\Phi^{-1}[1 - F_2(x)] \\ &+ r_X(\Phi^{-1}[1 - F_2(x)])^2)/2(1 - r_X^2)\}/(1 - r_X^2)^{1/2}, \end{aligned} \quad (1)$$

where r_X is the correlation between the experts' judgments. As in appendix B, the experts' medians m_1 and m_2 are taken as forecasts of X , and r_X is the correlation between the experts' forecast errors.

Using (1) requires specifying correlation r_X . In the context of combining expert probability distributions, it is rare that data are available, and so the correlation must be subjectively assessed. For this application, we used the "conditional fractile" assessment method as described

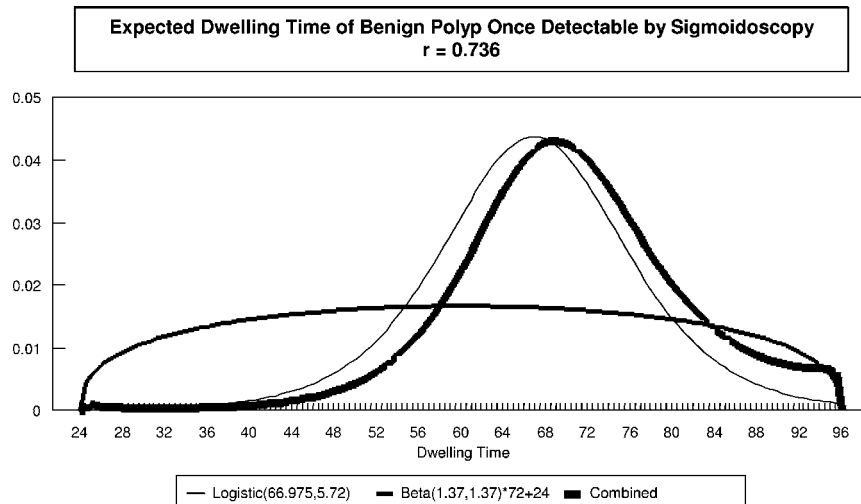


Figure 1. Individual and combined densities for X_7 , expected dwelling time for invasive procedure.

Table 3
Correlations for expert assessments of eight model variables.

Variable	Correlation
Sensitivity of FOBT	0.865
Proportion with polyps at age 50	0.845
Proportion of CRCs from polyps	0.901
Proportion of discovered early-stage CRCs	0.833
Expected dwelling time for early stage CRC	0.804
Expected dwelling time for late stage CRC	0.770
Expected dwelling time for invasive procedure	0.736
Expected dwelling time for FOBT	0.734

in [15,16]. Because Lacke directed the elicitation interviews with the experts, we believed that he was best suited to assess the correlation between the experts' judgments as required by Jouini and Clemen's aggregation model. Thus, with Clemen's assistance, Lacke assessed the correlations, as shown in table 3, as part of the model-building process. At the end of section 4, we report a sensitivity analysis of these correlations with respect to the final results of the model.

Given the eight correlations, we can now create the copula models for combining each pair of distributions. Here we describe the model for X_7 , the expected dwelling time for an invasive procedure. Recall that the densities elicited from the two experts were fitted as $f_1(X_7) = \text{Logistic}(66.975, 5.72)$ and $f_2(X_7) = \text{Beta}(1.37, 1.37) \cdot 72 + 24$. Applying the copula model (1), the combined density $f(X_7|f_1, f_2)$ was calculated at enough points to produce a relatively smooth curve. Figure 1 presents a graphical comparison of the univariate models and the copula model. Similar combination procedures were followed for the remaining seven variables.

Note that the mean of the combined density in figure 1 appears to exceed the means of both of the component densities. This is a phenomenon that can occur with combination models when the dependence level is relatively high. The canonical example is to imagine two forecasts h_1 and h_2 . The two forecasts are unbiased (i.e., their expected forecast errors are zero), and h_2 is known to be less accurate than h_1 (i.e., larger error variance). Now suppose that the two fore-

casts errors are perfectly correlated. Knowing h_1 and h_2 , and that $h_1 > h_2$, the combined forecast would exceed h_1 .

Although our model incorporates dependence among the expert probability distributions for the purpose of combining the distributions, we do not incorporate dependence among the uncertain model parameters. For example, it is conceivable that early and late stage expected dwelling times could be negatively correlated, depending on the definition of the transition form one stage to another. We chose not to incorporate these correlations for two reasons. First and foremost, very little information exists that would be helpful for the experts to assess such correlations, and developing such information was beyond the scope of the project. Second, Smith et al. [15] describe circumstances under which correlations can be ignored, including weak correlations among the variables; we doubt that the correlations among the parameters would be large enough to have a meaningful effect on the results. We return to this issue in the final section.

4. Model analysis and results

4.1. Simulation results

With the combined probability distributions in place, our enhanced version of Wagner's model was analyzed using Monte Carlo simulation. The simulation was performed in @RISK for 2000 iterations. The same seed (917) was used to allow for comparison across screening regimens.

Figure 2 graphs the means for CPLYs and CRCs prevented. Preferred screening procedures (according to the expected values) lie in the northwest direction in figure 2. Thus, regimens DCBE-3, CSCPE-5, CSCPE-5/FOBT, CSCPE-3, and CSCPE-3/FOBT create a frontier of non-dominated screening policies, for which no preference ordering can be stated at this point. This analysis does not include information regarding variability of CRCs prevented and CPLYs, however, which could affect the preference ordering of the regimens. A more extensive discussion of the simulation results can be found in [7].

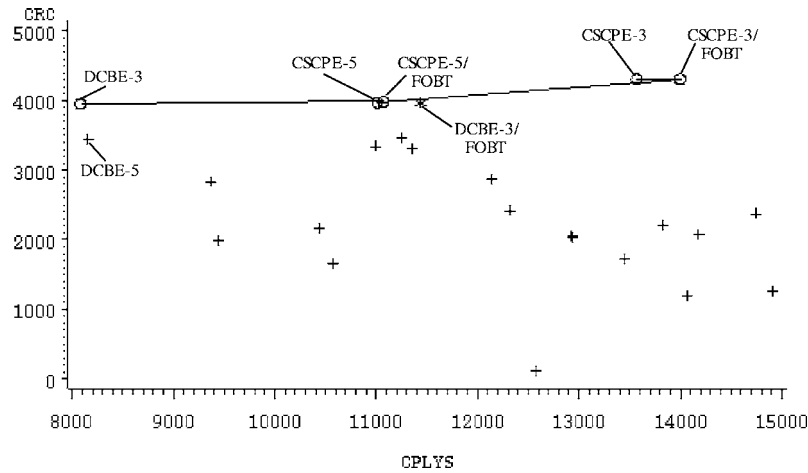


Figure 2. Means of CPLYs and CRCs prevented for colorectal screening procedures.

4.2. Utility analysis

Following standard decision analysis practices, a utility analysis was performed to provide a complete ranking of the regimens. Because the study was performed in a public health framework, we elicited utilities from Dr. Joseph Holliday, Director of the North Carolina Breast and Cervical Cancer Control Program (BCCCP), which is a public screening program with goals comparable to a colorectal screening program.

In the process of eliciting utilities, it was determined that mutual utility independence [18] was appropriate when CPLYs was partitioned into two ranges, less than \$50,000 and at least \$50,000. Dr. Holliday’s responses displayed risk aversion with respect to the number of CRCs prevented and to CPLYs in the \$0–\$50,000 range. However, Dr. Holliday’s responses were risk seeking for CPLYs in the \$50,000–\$125,000 range. For convenience, we rescaled CPLYs to represent “Savings” so that savings (Y) = \$50,000 – CPLYs. Thus, savings = – \$75,000 corresponds to CPLYs = \$125,000 and savings = \$50,000 corresponds to CPLYs = \$0.

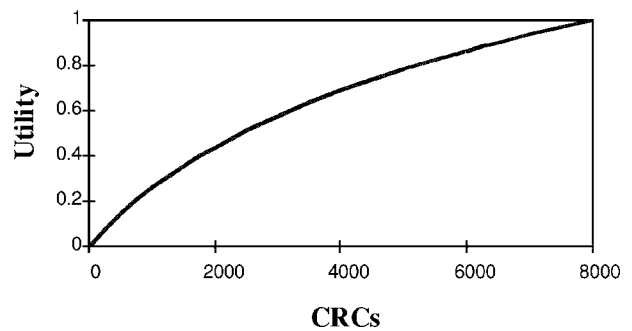
In fitting the individual utility functions, logarithmic models were chosen to represent risk aversion, while an exponential model was used to represent the risk-seeking portion of the utility function for savings. Specifically, the utility functions are

$$\begin{aligned}
 u_X(x) &= -4.423 + 0.590 \ln(x + 1800), \\
 u_Y(y) &= \begin{cases} 0.5 \exp((-0.000000277)y^2), & y \leq 0, \\ -6.034 + 0.617 \ln(y + 40000), & y > 0, \end{cases} \quad (2)
 \end{aligned}$$

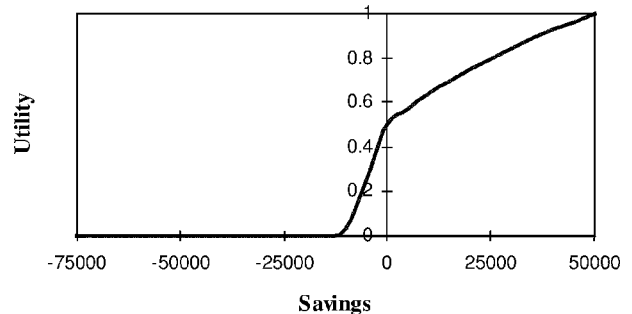
where X is the number of CRCs prevented and Y is savings. The constants are chosen so that the utility functions range from 0 to 1. The graphs of these functions are shown in figure 3.

Since we were able to assume mutual utility independence, the functional form of the two-attribute utility function is

$$u(x, y) = k_X u_X(x) + k_Y u_Y(y) + k_{XY} u_X(x) u_Y(y), \quad (3)$$



(a)



(b)

Figure 3. Utility functions for CRCs prevented (a) and savings (b).

where

$$k_X + k_Y + k_{XY} = 1.$$

Based on Holliday’s trade-off assessments, the attribute weights are

$$k_X = 0.7, \quad k_Y = 0.183, \quad k_{XY} = 0.117.$$

The relatively large value of k_X indicates that CRCs prevented is the more important attribute, and the positive value for k_{XY} suggests that CRCs prevented and savings are complementary. Our interpretation is that a high level for savings, when coupled with a high level for CRCs prevented,

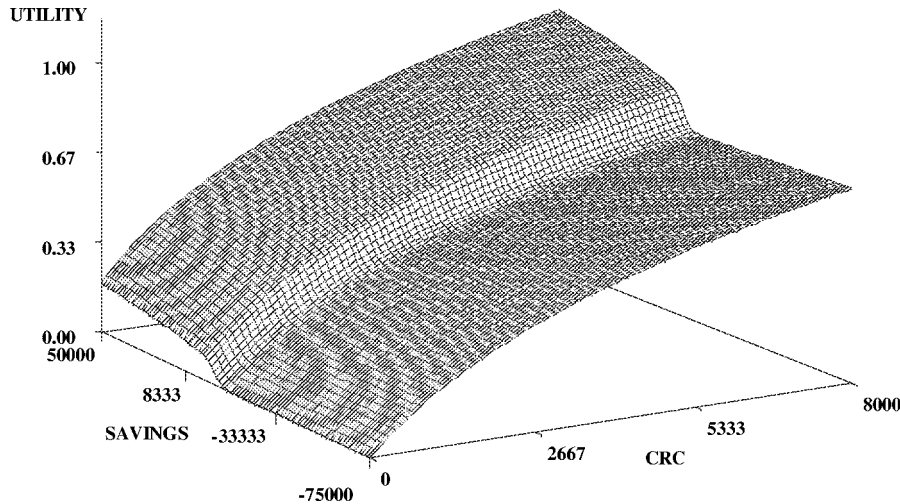


Figure 4. Graph of two-attribute utility function for number of CRCs prevented (X) and savings ($Y = \$50,000 - \text{CPLYS}$).

Table 4
Expected utilities for 25 screening regimens.

Regimen	Expected utility	Regimen	Expected utility
CSCPE-3	0.745	CSCPE-10	0.559
CSCPE-3/FOBT	0.744	FSIG-3	0.544
DCBE-3	0.728	FSIG-3/FOBT	0.544
CSCPE-5/FOBT	0.724	DCBE-10/FOBT	0.526
CSCPE-5	0.723	FSIG-5/FOBT	0.525
DCBE-3/FOBT	0.722	FSIG-5	0.525
DCBE-5	0.682	DCBE-10	0.521
DCBE-5/FOBT	0.678	FSIG-7/FOBT	0.482
CSCPE-7/FOBT	0.666	FSIG-7	0.479
CSCPE-7	0.663	FSIG-10/FOBT	0.412
DCBE-7	0.619	FSIG-10	0.403
DCBE-7/FOBT	0.618	FOBT	0.184
CSCPE-10/FOBT	0.569		

Table 5
Attribute weight sensitivity analysis on the utility rank order of top six regimens, along with CSCPE-7/FOBT and CSCPE-7.

	Value of k_X^a				Baseline ^b
	0.1	0.2	0.3	0.4	
CSCPE-3	5	5	2	2	1
CSCPE-3/FOBT	6	6	5	3	2
DCBE-3	1	1	1	1	3
CSCPE-5/FOBT	3	2	3	4	4
CSCPE-5	2	3	4	5	5
DCBE-3/FOBT	4	4	6	6	6

^a The columns beneath the values of k_X indicate the regimen's rank, given the indicated value of k_X . Note that $k_Y = 0.883 - k_X$ and $k_{XY} = 0.117$. As k_X increases, the relative importance of CRCs prevented also increases.

^b The baseline rank order of the screening regimens is preserved for $k_X > 0.475$.

would be viewed as being more valuable than the same amount of savings when fewer CRCs are prevented.

Substituting for $u_X(x)$, $u_Y(y)$, k_X , k_Y , and k_{XY} in (3), the expression for the two-attribute utility function is

$$u(x, y) = \begin{cases} 0.413 \ln(x + 1800) - 0.169 \exp(-0.000000277y^2) + 0.035 \ln(x + 1800) \exp(-0.000000277y^2) - 3.096, & y \leq 0, \\ 0.043 \ln(x + 1800) \ln(y + 40000) - 0.208 \ln(y + 40000) - 0.005 \ln(x + 1800) - 1.0620, & 0 < y \leq 50000, \\ 0.413 \ln(x + 1800) - 3.096, & y > 50000, \end{cases} \quad (4)$$

which is displayed in figure 4.

In the Monte Carlo simulation procedure, $u(x, y)$ was calculated for each iteration, and the mean of these was taken as the expected utility. The results are shown in table 4. The top five regimens are the same five that constituted the frontier in the graphical analysis, followed again by DCBE-3/FOBT.

4.3. Sensitivity analyses

The first sensitivity analysis focuses on the diagnosticity of DCBE. In the original model [3], and hence in our baseline analysis, DCBE's sensitivity (i.e., the probability of a true positive result) was assumed to be 0.8. Rex et al. [19] show results supporting this value. Winawer et al. [20], however, suggest that the sensitivity might be much lower, possibly lower than 0.5. Using this value in the model, DCBE-3 and DCBE-3/FOBT drop to seventh and eighth, with CSCPE-7/FOBT and CSCPE-7 taking over fifth and sixth place, respectively.

Our second sensitivity analysis focused on the utility-function weights. We varied k_X and k_Y while keeping $k_X + k_Y = 0.883$ and $k_{XY} = 0.117$. Thus, this sensitivity analysis retains the complementarity of the two attributes but varies their relative importance. The results are displayed in table 5.

The rank order of the regimens does not change until k_X falls below 0.475. At this value, DCBE-3 becomes the top-ranking regimen. At $k_X = 0.3$, CSCPE-3/FOBT falls in rank from three to five. Additional decreases in k_X result in further rank reductions of CSCPE-3 and CSCPE-3/FOBT.

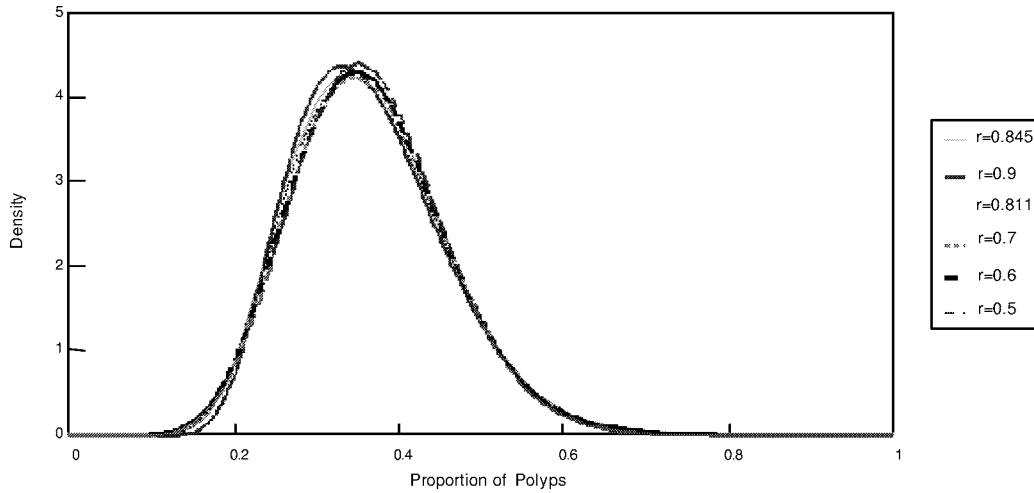


Figure 5. Correlation sensitivity analysis on combined density for proportion of polyps at age 50. For various correlations.

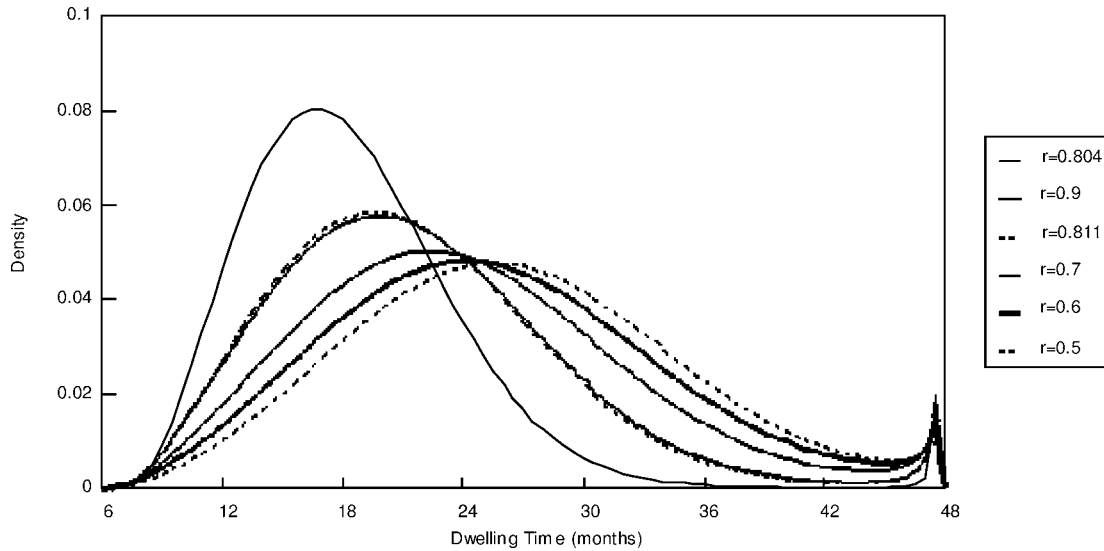


Figure 6. Correlation sensitivity analysis on combined density for early-stage expected dwelling time. For various correlations.

These changes reflect the debate over the quality of preventive measures versus the cost. Holliday’s assessments indicated a high weight for CRCs prevented relative to CPLYs. As the relative importance of CPLYs increases, preference may be given to less costly regimens that have lower sensitivity or less frequent screenings.

If we use the more conservative assumption that DCBE sensitivity = 0.50, though, the rank changes reported above become irrelevant. Under this assumption, the DCBE regimens have rankings lower than all of the CSCPE regimens shown in table 5, and CSCPE-5 becomes preferable for low values of k_X . (Podolsky [21] suggests that a five-year interval for colonoscopy may be appropriate.)

In the third sensitivity analysis, we recalculated the aggregated distributions five times using correlations of 0.9, 0.811 (the average of the eight elicited correlations), 0.7, 0.6, and 0.5. In each round, the same correlation was used for each of the eight variables. This range is consistent with [16], who report average absolute errors of about 0.25

among individuals using the conditional fractile assessment method.

Changing the correlation can have relatively little effect on the combined density, or it can have a strong effect. Figure 5 shows the combined density of proportion of polyps at age 50 under the various correlations; it is obvious that changing the correlation has very little effect on the combined density. In this case, the individual expert densities are similar in shape, and so this result is not surprising. Figure 6, though, shows that the combined density for early-stage expected dwelling time can change dramatically with changes in the correlation.

Given that the aggregate densities can be sensitive to the assessed correlation in the aggregation model, is the rank ordering of the screening regimens similarly sensitive? The only change that occurs is between CSCPE-5 and DCBE-3/FOBT, which rank 5th and 6th, respectively, in the baseline analysis, but switch places for correlations below 0.7. Thus, despite changes in the aggregated densities, the lack

of a high level of accuracy in assessing correlations does not substantially affect on the final ranking of the top screening regimens.

A variety of additional sensitivity analyses were performed, all of which confirmed the robustness of the original rankings to changes in various model parameters. Details are available from the authors.

5. Discussion and conclusion

Our analysis shows that the optimal screening regimen for asymptomatic, average-risk individuals, aged 50–85, is colonoscopy every three years. Screening regimens utilizing a colonoscopy every five years, or a barium enema every three years, with or without an annual FOBT, have expected utilities very close to the expected utility for CSCPE-3. Our results are robust to changes in the assessed correlation between experts, as well as to the values of the trade-off weights in the two-attribute utility function. Changes in the utility function for savings could, however, result in a switch between DCBE-3 and CSCPE-3, due to the substantially lower cost associated with DCBE.

Our results are consistent with those reported in [22] based on a deterministic model. In contrast, our model incorporates uncertainty in the form of expert judgments, uses Monte Carlo simulation to analyze various screening policies, and measures their value with a two-attribute utility function for CRCs prevented and cost per life-year saved. Risk associated with adenomatous polyps is also reported in [23], concluding that polyps in the distal colon are more likely to develop into CRCs. In our model, we assumed the same risk for polyps at any location; the greater risk for distal polyps, though, only increases the effectiveness and value of colonoscopy. Using a more conservative estimate of DCBE sensitivity only lowers DCBE’s effectiveness and value relative to CSCPE-3.

There are two specific ways in which further sensitivity analysis could enhance the study we report here. First, we chose to follow the standard decision-analysis approach of eliciting and using preferences in a utility analysis in order to find a rank ordering of the screening regimens. An alternative approach would be to perform a sensitivity analysis to determine the optimal regimen as a function of the unknown parameters. Such an analysis could indicate which parameters influence the choice of regimen the most, and hence could help focus research on learning more about the most crucial parameters.

Second, as discussed above, we chose not to incorporate correlations among the uncertain parameters on the model. Although we believe that correlations among the parameters would be too weak to have a substantial impact on the results, a sensitivity analysis of parameter correlations could illuminate this issue and indicate whether further research effort to assess and incorporate those correlations would be fruitful.

Many decision and risk analyses rely on expert judgment, and researchers often must reconcile judgments from multi-

ple experts. Our study relies on the judgments from two knowledgeable experts, and although we followed standard practice for eliciting and incorporating expert judgments, our reliance on only two experts may be seen as a limitation of the study. Part of the methodological contribution of this article is demonstrating the use of Jouini and Clemen’s aggregation method with the multivariate normal copula. The calculations can be performed in a spreadsheet or statistical analysis program. The sensitivity analysis suggests that correlations for the aggregation model need not be assessed with great precision, as also suggested by [15].

Acknowledgements

The authors are grateful to Tim Aldrich, Yayha Fathi, Harvey Gold, Joseph Holliday, Joseph Lipscomb, John Monahan, Dawn Provenzale, Robert Sandler, Charles E. Smith, and Judith L. Wagner for their assistance with the work leading to this paper. In addition, we gratefully acknowledge support from The North Carolina Central Cancer Registry and from the National Science Foundation under Grant SES-0084383.

Appendix A. Parameterizations for density functions with formulae for calculating means and standard deviations [11]

Beta(α, β) (written as Beta(α, β)($d - c$) + c for $c \leq x \leq d$)

$$f(x | \alpha, \beta) = \begin{cases} \left(\frac{1}{d-c}\right)^{\alpha+\beta-1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} (x-c)^{\alpha-1} (d-x)^{\beta-1}, & c \leq x \leq d, \\ 0, & \text{elsewhere,} \end{cases}$$

$$E(X) = \frac{\alpha}{\alpha+\beta}(d-c) + c,$$

$$\text{s.d.}(X) = \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}}.$$

Gamma(α, β)

$$f(x | \alpha, \beta) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, & x \geq 0, \alpha > 0, \beta > 0, \\ 0, & \text{elsewhere,} \end{cases}$$

$$E(X) = \alpha\beta, \quad \text{s.d.}(X) = \sqrt{\alpha\beta^2}.$$

Logistic(α, β)

$$f(x | \alpha, \beta) = \begin{cases} \frac{\exp\{-(x-\alpha)/\beta\}}{\beta[1+\exp\{-(x-\alpha)/\beta\}]^2}, & -\infty < x < \infty, -\infty < \mu < \infty, \beta > 0, \\ 0, & \text{elsewhere,} \end{cases}$$

$$E(X) = \alpha, \quad \text{s.d.}(X) = \frac{\pi\beta}{\sqrt{3}}.$$

Lognormal(μ, σ)

$$f(x | \mu, \sigma) = \begin{cases} \frac{1}{x\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{(\ln x - \mu_1)^2}{2\sigma_1^2}\right\}, \\ x > 0, \mu > 0, \sigma > 0, \\ 0, & \text{elsewhere,} \end{cases}$$

where

$$\mu_1 = \ln\left[\frac{\mu^2}{\sqrt{\sigma^2 + \mu^2}}\right], \quad \sigma_1 = \sqrt{\ln\left[\frac{\sigma^2 + \mu^2}{\mu^2}\right]},$$

$$E(X) = \mu, \quad \text{s.d.}(X) = \sigma.$$

Weibull(α, β)

$$f(x | \alpha, \beta) = \begin{cases} \alpha\beta^{-\alpha}x^{\alpha-1} \exp\left\{-\left(\frac{x}{\beta}\right)^\alpha\right\}, \\ x > 0, \alpha > 0, \beta > 0, \\ 0, & \text{elsewhere,} \end{cases}$$

$$E(X) = \frac{\beta}{\alpha} \Gamma\left(\frac{1}{\alpha}\right),$$

$$\text{s.d.}(X) = \beta \sqrt{\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)}.$$

Appendix B. Overview of copulas and description of Jouini and Clemen's [4] model for combining expert probability distributions

B.1. Overview of copulas

Suppose that two continuous random variables, X and Y , have differentiable marginal distribution functions $F_X(x)$ and $F_Y(y)$ and corresponding marginal densities $f_X(x)$ and $f_Y(y)$. If these two random variables are independent, their joint density $f_{XY}(x, y)$ can be written as the product $f_X(x)f_Y(y)$. The notion of a copula allows us to write the joint distribution as a function of the marginals even though X and Y may not be independent. In particular, Sklar's Theorem [24] says that any joint distribution can be written as a function of the marginal distribution functions. X and Y 's joint cumulative distribution function can be written as $F_{XY}(x, y) = C[F_X(x), F_Y(y)]$, where $C[u, v]$ is a joint distribution with uniform marginals. More helpful, perhaps, is that the corresponding joint density is

$$f_{XY}(x, y) = f_X(x)f_Y(y)c[F_X(x), F_Y(y)], \quad (\text{B1})$$

where $c[u, v] = \partial^2 C[u, v]/\partial u \partial v$. The function $c[\]$ is sometimes called the "copula density." Thus, paralleling the independent case, the joint density f_{XY} is the product of the marginal densities and the copula density. The copula density encodes information about the dependence between variables X and Y .

Sklar's theorem shows that, in principle, any joint distribution can be written in copula form. In practice, however, the modeler does not derive the copula from the joint distribution; rather, it is more common to choose a specific

member from a family of copulas as a basis for modeling the dependence among the variables. Many families of copulas are available for representing dependence, and Frees and Valdez [16] describe several. Clemen and Reilly [18] show how to derive and use the copula that underlies the multivariate normal density for modeling purposes in decision and risk analysis. Using the bivariate normal copula, the joint density for X and Y can be written as

$$f_{XY}(x, y | r) = f_X(x) \times f_Y(y) \\ \times \exp\{-r(r(\Phi^{-1}[F_X(x)])^2 \\ - 2\Phi^{-1}[F_X(x)]\Phi^{-1}[F_Y(y)] \\ + r(\Phi^{-1}[F_Y(y)]^2)/2(1-r^2)/(1-r^2)^{1/2}\}, \quad (\text{B2})$$

where $\Phi^{-1}[\]$ is the inverse of the standard normal cumulative distribution function, and r is the correlation between X and Y . Clemen and Reilly also discuss some ways to assess the correlation r ; for further evaluation of correlation-assessment methods, see [18]. Equation (B2) fully specifies the joint density for X and Y , which can then be used in an analytical model. We use it below in a model for aggregating the experts' probabilities in the CRC model. For more information on copulas, see [15,25,26].

B.2. Jouini and Clemen's copula-based combination model

We return now to the problem of combining our experts' probability distributions using Jouini and Clemen's copula-based approach. Using our notation from above, two experts have each assessed a cumulative distribution function (CDF) for variable X , yielding $F_1(x)$ and $F_2(x)$. As before, assume that these CDFs are differentiable and have corresponding densities $f_1(x)$ and $f_2(x)$.

Modeling dependence among experts' probability distributions can, in principle, be quite complicated. The analyst could potentially model all aspects of dependence among all of the fractiles of one expert's distribution with those of the other expert. Even if the model is limited to a finite number of fractiles, it still would be a complex multivariate model for which no truly general, operational approach has been devised. Jouini and Clemen invoke some relatively simple assumptions, however, in order to construct a copula-based aggregation model.

The first assumption is that, if the DM were to consult only one expert, the DM would adopt the expert's reported distribution as his or her own. Thus, if only expert i were consulted, $f_i(x)$ would become the DM's posterior density for X : $f_{\text{DM}}(x|f_i) = f_i(x)$.

From this, we can use Bayes' Theorem (assuming a constant and possibly improper prior) to infer the DM's likelihood function

$$l(f_i | x) = f_i(x). \quad (\text{B3})$$

The second assumption is that the medians (m_i) of the experts' distributions serve as forecasts of X , and that dependence among the experts can be thought of in terms of

the dependence among their medians relative to a realization of X . For example, if one expert reports a median that turns out to be greater than X , then a positively-dependent second expert would tend also to report a median that is too high. This is a relatively simple way to think about dependence among the experts, but it is consistent with an intuitive notion of how experts might be dependent in terms of accuracy.

To construct the dependence model, Jouini and Clemen consider the experts' medians as not-yet-reported random variables for which conditional distributions can be derived, given x . The conditional distribution function for m_i is given by

$$L_i(m_i | x) = 1 - F_i(c_i + x - m_i), \tag{B4}$$

where c_i is a constant that must be equal to the median of f_i . The requirement that c_i equal the median of f_i is required in order to satisfy the first assumption above.

We now have everything necessary to write the multivariate likelihood function for the aggregation model. Write the joint likelihood function using a copula representation as in (B1):

$$\begin{aligned} l(f_1, f_2 | x) &= l_1(f_1 | x)l_2(f_2 | x)c[L_1(f_1 | x), L_2(f_2 | x)] \\ &= f_1(x)f_2(x)c[L_1(f_1 | x), L_2(f_2 | x)], \end{aligned} \tag{B5}$$

where the second line follows from (B3). Then complete the dependence model by assuming that $L_i(f_i | x) = L_i(m_i | x) = 1 - F_i(c_i + x - m_i)$, which is consistent with our assumption that dependence is modeled only through the medians of the experts' distributions. (Note: treating (B5) rigorously requires that the arguments L_i of the copula density correspond to the univariate functions l_i . In their derivation, Jouini and Clemen show that conditional densities l_i corresponding to the distributions L_i in (B4) ultimately reduce to f_i in the analysis of (B5). For details, see [4].)

At this point, the likelihood function is fully specified. Invoking the assumption of a non-informative, constant (and possibly improper) prior distribution, applying Bayes' theorem amounts to normalizing the joint likelihood function. Noting that the posterior distribution $f_{DM}(x | f_1, f_2)$ involves fixed values for the medians, we can write $L_i(m_i | x) = 1 - F_i(c_i + x - m_i) = 1 - F_i(x)$ to obtain

$$f_{DM}(x | f_1, f_2) = f_1(x)f_2(x)c[1 - F_1(x), 1 - F_2(x)], \tag{B6}$$

which is Jouini and Clemen's result.

References

[1] *Cancer Facts and Figures 2001* (American Cancer Society, Atlanta, 2001).
 [2] L. Ries, C. Kosary, B. Hankey, B. Miller, L. Clegg and B. Edwards, eds., *SEER Cancer Statistics Review, 1973–1996* (National Cancer Institute, Bethesda, MD, NIH Publication 99-2789, 1999).

[3] J. Wagner, S. Tunis, M. Brown, A. Ching and R. Almeida, Cost effectiveness of colorectal cancer screening in average-risk adults, in: *Prevention and Early Detection of Colorectal Cancer*, eds. G. Young, P. Rozen and B. Levin (WB Saunders, Philadelphia, 1996).
 [4] M. Jouini and R. Clemen, Copula models for aggregating expert opinions, *Operations Research* 44 (1996) 444–457.
 [5] J. Lipscomb, M. Weinstein and G. Torrance, Time preference, in: *Cost-Effectiveness in Health and Medicine*, eds. M. Gold, J. Siegel, L. Russell and M. Weinstein (Oxford University Press, New York, 1996).
 [6] Medicode, Inc., Salt Lake City, UT (1997).
 [7] C. Lacke, Decision analytic modeling of colorectal cancer screening Policies, Ph.D. Dissertation, North Carolina State University (1998).
 [8] R. Clemen, *Making Hard Decisions: An Introduction to Decision Analysis*, 2nd edn. (Duxbury Press, Pacific Grove, CA, 1996).
 [9] M. Meyer and J. Booker, *Eliciting and Analyzing Expert Judgement: A Practical Guide* (Academic Press, London, 1991).
 [10] M. Morgan and M. Henrion, *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis* (Cambridge University Press, Cambridge, 1990).
 [11] *BestFit User's Guide* (Palisade Corporation, Decker, NY, 1996).
 [12] S. French, Group consensus probability distributions: A critical survey, in: *Bayesian Statistics*, Vol. 2, eds. J. Bernardo, M. DeGroot, D. Lindley and A. Smith (North-Holland, Amsterdam, 1985).
 [13] C. Genest and J. Zidek, Combining probability distributions: A critique and annotated bibliography, *Statistical Science* 1 (1986) 114–135.
 [14] R. Clemen and R. Winkler, Combining probability distributions from experts in risk analysis, *Risk Analysis* 19 (1999) 187–203.
 [15] A. Smith, P. Ryan and J. Evans, The effect of neglecting correlations when propagating uncertainty and estimating the population distribution of risk, *Risk Analysis* 12 (1992) 467–474.
 [16] E. Frees and E. Valdez, Understanding relationships using copulas, *North American Actuarial Journal* 2 (1998) 1–25.
 [17] R. Nelsen, *An Introduction to Copulas* (Springer, New York, 1999).
 [18] R. Clemen and T. Reilly, Correlations and copulas for decision and risk analysis, *Management Science* 45 (1999) 208–224.
 [19] A. Sklar, Fonctions de répartition à n dimensions et leurs marges, *Publications de l'Institut Statistique de l'Université de Paris* 8 (1959) 229–231.
 [20] R. Clemen, G. Fischer and R. Winkler, Assessing dependence: Some experimental results, *Management Science*, in press.
 [21] R. Keeney and H. Raiffa, *Decisions with Multiple Objectives: Preference and Value Tradeoffs* (Wiley, New York, 1976).
 [22] D. Rex, E. Rahmani, J. Haseman, G. Lemmel, S. Kaster and J. Buckley, Relative sensitivity of colonoscopy and barium enema for detection of colorectal cancer in clinical practice, *Gastroenterology* 112(1) (January, 1997) 17–23.
 [23] S. Winawer, E. Stewart and A. Zauber, A comparison of colonoscopy and double contrast barium enema for surveillance after polypectomy, *New England Journal of Medicine* 342(24) (June 2000) 1766–1772.
 [24] D. Podolsky, Going the distance – the case for true colorectal-cancer screening, *New England Journal of Medicine* 343(3) (July 2000) 207–208.
 [25] D. Lieberman, D. Weiss, J. Bond, D. Ahnen, H. Garewal and G. Chejfec, Use of colonoscopy to screen asymptomatic adults for colorectal cancer, *New England Journal of Medicine* 343(3) (July 2000) 162–168.
 [26] T. Imperiale, D. Wagner, C. Lin, G. Larkin, J. Rogge and D. Ransohoff, Risk of advanced proximal neoplasms in asymptomatic adults according to the distal colorectal findings, *New England Journal of Medicine* 343(3) (July 2000) 169–174.