

IMPORTANT FACTORS IN SCREENING FOR COLORECTAL CANCER

Reza Yaesoubi

Edward P. Fitts Department of
Industrial and Systems Engineering
400 Daniels Hall,
North Carolina State University
Raleigh, N.C. 27695, U.S.A.

Stephen D. Roberts

Edward P. Fitts Department of
Industrial and Systems Engineering
400 Daniels Hall,
North Carolina State University
Raleigh, N.C. 27695, U.S.A.

ABSTRACT

A complex, stochastic simulation model of Colorectal Cancer (CRC) is examined through factor screening to determine which factors in the model are important. The factor screening employs a Resolution IV 2^k Fractional Factorial experimental design. The factors are examined in terms of their impact on cost, quality-adjusted life-years (QALY), and cost per QALYs. Out of 72 factors, eight factors were determined to be most important and observed as “driving factors” in the CRC model. Surprisingly these factors were consistently important for all outcomes. However the limitations of the experimental design may have constrained the important factors to factors related only to the natural history of the disease and therefore subject to minimal control.

1 INTRODUCTION

A medical simulation model called the Vanderbilt/NC State model is a stochastic, discrete-event simulation model of the natural history of Colorectal Cancer (CRC) (Roberts et al. 2008). This model simulates a population over time which may include a mixture of patients with different birth years, races, genders, and family histories of colorectal neoplasia. The model has been developed using an object-oriented simulation platform driven by an independent database to provide a complete representation of the potential stochastic impact of CRC. The discrete-event representation of the natural history models changes in the CRC state of an individual throughout his/her natural lifetime. Screening can intervene in the CRC process by detecting adenomas and early cancers. Removing these neoplasia changes the future outcomes, thus potentially extending life. This model produces discounted costs and quality-adjusted life-years (QALYs) for screening decisions as the primary outcomes although various natural history data may also be collected and reported. The model has been carefully verified, calibrated,

and validated and used to determine the cost-effectiveness of CRC screening alternatives (Tafazzoli et al. 2005).

Complex medical decision-making simulation models such as the CRC model often consist of a large number of *inputs* which are usually referred to as *factors* in Design of Experiments (DOE). The output (response) of a simulation model can be explored by using designed experiments. It is, however, very common that due to their complexity, studying the behavior of the large-scale medical simulation models become prohibitively costly and requiring time-consuming experimental design analyses. *Screening experiments*, which assume that only a few factors are really important (*parsimony principle*), examine many factors with the objective being to identify those factors (if any) that have significant (important) effects on a selected response.

Screening is generally employed in the *pilot phase* of complex simulation studies (Bettonvil et al. 1996). The factors identified as being important can be further explored in later phases; e.g. the important factors might be cast as a metamodel and used in optimization. Moreover, the results of factor screening can be used not only for confirming prior expectations (which is as an important step in validating the simulation model), but they are also informative when the simulation provides insights that do not match expectations; for example, it is possible that a factor believed to be important by the content experts turns out to be statistically insignificant or unimportant.

This paper presents the results of applying factor screening methods to the Colorectal Cancer (CRC) Simulation model. Many of the input variables for this model are uncertain and modeled with random variables. Most of the input factors correspond to *distribution-based variables* and the rest are *constant factors*. Distribution-based variables are those whose values are determined according to a distribution function whose parameters are referred to as *factors*. For example, in the CRC Model variable-1 follows a Johnson S_B distribution with factors 1-4 as distribution parameters: delta, gamma, minimum, and

maximum, respectively. In almost all cases, the factors associated with a common distribution variable should be confounded, i.e. their values cannot change independently. After confounding the necessary factors, the factor screening experiment is performed on 72 factors.

2 LITERATURE REVIEW

Among the most widely used factor screening methods are 2^k Fractional Factorial (2^k FF) designs (Montgomery 2001). These designs assume that certain higher-order interactions are negligible so they can be confounded with main or second-order interaction effects. Also, it is assumed that errors are approximately normally distributed with mean zero and a *homogenous* variance σ^2 . One of the advantages of 2^k FF designs is that they are very easy to construct and already available in almost all statistical software. Factorial designs have several attractive properties. Since more than one factor can be examined at a time, these designs can identify important interaction effects. They are also *orthogonal* designs, i.e. the pair-wise correlation between any two columns (factors) in the design matrix is equal to zero. This simplifies the analysis of the fitted model, because estimates of the factors' effects (β_i 's) and their contribution to the explanatory power of the regression metamodel will not depend on what other regressors are included in the regression model. Although ordinary least squares regression assumes that the error term is identically distributed but the regression coefficients are still unbiased estimators even if the underlying variance is not constant.

In 2^k FF designs factors are examined at only two levels; examining each factor at only two levels (the low and high values) does not reveal how the simulation output behaves for the factor combinations in the interior of the experimental region. Moreover, it is possible that the choice of low and high level for factors cancels the interaction (Trocine et al. 2000). One of the other major concerns about the fractional factorial design is that this design may confound the interaction of significant importance with other effects; thus the result is combined with other main effects or interactions and nothing can be determined about the individual interactions within this confounded structure. Another potential concern with the use of two-level factorial designs is the assumption of linearity in the factor effects. A simple and highly effective solution to this problem is to augment the 2^k design with *axial runs*. The resulting design, called a *Central Composite Design* (Montgomery 2001), can now be used to fit the second-order metamodel. When the number of factors is fairly large (e.g. more than 10), Central Composite Designs require a large number of design points which deteriorate the efficiency of the design.

Sequential Bifurcation, introduced by Bettonvil (1996), was mainly designed for deterministic simulation

models and was developed based the Group Screening designs: (Johnson et al. 1999), (Kleijnen 1987), (Li 1962), (Ottieno et al. 1984), (Patal et al. 1984), (Patel 1962). SB, like all other group screening techniques, assumes a low-order polynomial metamodel for the response function of the simulation model, and known signs for the main effects in order to assure that individual effects do not compensate for the effects of each other within a group. The criterion that SB uses for declaring a factor as important is the absolute value of the factor's main effect.

Cheng (1997) extended the SB to handle stochastic simulations where the response is stochastic and subject to significant error. First-order polynomial (main effect only) for the response function and a homogenous variance over the entire experimental region are assumed. Cheng's method also assumes known signs of main effects. The method requires two parameters from user: delta limit (δ), which is the value that the main effect of a factor should reach to be considered important, and alpha (α), which is the probability of declaring an important factor unimportant.

Cheng's method does not provide performance guarantee in terms of probability of classifying the factors correctly. Controlled Sequential Bifurcation (CSB), proposed by Wan et al. (2003) is a procedure that incorporates a two-stage hypothesis-testing approach into SB to control the power (i.e., the probability that an important effect is correctly classified) at each bifurcation step and Type I Error (i.e., the probability an effect is classified as important when it is not) for each factor under *heterogeneous* variance conditions. Similar to basic SB, CSB assumes a first-order polynomial function for the response and known direction for main effects. For CSB, the analyst must specify two thresholds: Δ_0 and Δ_1 . The lower threshold (Δ_0) indicates the level that the main effect of a factor must reach to be considered *important*; and if the main effect of a factor is larger than the higher threshold (Δ_1), the factor is considered *critical*; and otherwise if $\Delta_0 < \beta_j < \Delta_1$, it is classified as being important. More specifically, for those factors with effects $\leq \Delta_0$, CSB controls the Type I Error by declaring them important to be less than α ; and for those factors with effects $\geq \Delta_1$, CSB provides power for identifying them as important to be greater than γ . Those factors whose effects fall between Δ_0 and Δ_1 are considered important and the CSB procedure has reasonable, though not guaranteed, power to identify them (Wan et al. 2003). One of the key advantages of CSB is that the assumption of variance homogeneity and error independence is relaxed. Therefore the dependence of outputs across different factor settings due to common random numbers is permitted by CSB. For the special case, where $\alpha = 1 - \gamma$ (Type I error is equal to one minus power), Wan et al. (2005) implemented a fully sequential test in CSB that has the same error control as the two-stage testing procedure.

Wan et al. (2004) improved the CSB procedure by incorporating a fold-over design in the hypothesis test to identify important main effects even when two-factor interactions and quadratic terms are present. The new procedure, called CSB-X, still has the same error control for screening main effects. CSB-X is, however, not able to estimate interaction and quadratic effects.

The assumption of known direction for the main effects is not always realistic, especially when the simulation model is complex and includes many factors. (Sanchez et al. 2005) proposed a hybrid two-phase approach, FF-CSB, to relax this assumption. Phase 1 uses saturated or nearly-saturated fractional factorial experiment to estimate the signs and magnitudes of the effects. All the factors are then divided into positive and negative groups and sorted within each group factors. Next in Phase 2, original CSB is applied on the “positive” and “negative” groups separately. Sanchez et al. (2005) showed via numerical results that even with the extra effort in pre-screening, FF-CSB is generally more effective and efficient than CSB.

Another method which was proposed by Shen et al. (2005) to overcome the limitations of CSB is Controlled Sequential Factorial Design (CSFD). CSFD uses qualified hypothesis testing procedure with a sequential traditional factorial design to provide simultaneous Type I Error and power control for each interested effect. The first step of CSFD is to select a factorial design which will be sequentially implemented. For each effect, CSFD first computes an estimate of the effect coefficient from the available replications. Then, the estimate of the effect coefficient and its sample variance are computed; and finally, a hypothesis testing procedure will be used to classify the effect as important or unimportant, and determine if more replications are necessary. CSFD can work under heterogeneous variance conditions; also, this method can screen any main effects and interactions without assuming the directions of effects to be known. In addition, unlike CSB, in CSFD the generated observations in early screening processes can be utilized later to estimate the effects of interest.

Shen et al. (2006) proposed a hybrid method that combines CSB and CSFD to achieve better efficiency. The basic idea is to apply CSFD to screen those likely important effects (which is assumed to be small percentage of all effects) and CSB (or CSB-X) to screen those likely unimportant effects. The proposed hybrid method consists of two phases. In Phase 1, all the factors' main effects are estimated by using a fractional factorial design and each factor is explicitly assigned into one of the three groups as following: assign factors with potential important effects to group IMP; assign other factors to either group POS or NEG based on the directions of their estimated main effects. As pointed out by Shen et al. (2006), one of the drawbacks of this method is that if a factor asso-

ciated with a significant interaction effect is assigned to an unimportant group in Phase 1, there is no chance that this interaction could be classified as important in Phase 2 since CSB-X cannot screen interactions. In an empirical evaluation, Shen et al. (2006) showed that the proposed hybrid method is the most efficient one among CSFD and FF-CSB; however, like CSB, CSB-X, and FF-CSB, the efficiency of this method deteriorated significantly as the response variance increases.

3 FACTOR SCREENING THE CRC MODEL

3.1 The CRC Simulations

The CRC simulation modeled white males with no family history who began colonoscopy screening at age 50. The colonoscopy screening followed standard medical protocol with 10 year follow-up for negative findings. Positive screening results may be limited to removing the adenomatous polyp during screening with a three year follow-up. If the screening reveals colorectal cancer, then surgery is performed to remove the cancer. The extent of the surgery is determined by the state of the cancer and follow-up is appropriate to that state.

Compliance with screening is based on a general compliance that causes the individual to be never-compliant, one-time, or probabilistic compliant. The compliance affect first time screening as well as compliance to surveillance. The compliance variable specifications were part of the calibration used in validation.

In general, input variables for the CRC model were estimated through calibration from clinical trials and national data sources. Some of the data had to be interpreted by the physicians associated with the CRC model development and with a panel of selected medical experts. The final model was extensively tested and validated.

3.2 Initial Step for Factor Screening: Testing the Equality of Variance

Before performing factor screening on stochastic models, it is suggested that the experimental region be explored by using an inexpensive design such as small Latin Hypercube (LH) design or a Plackett-Burman (PB) design. Latin Hypercube Sampling (LHS) was proposed by McKay et al. (1979) for situations involving a relatively large number of factors. LH designs have good space-filling properties; that is, the design points are scattered throughout the experimental region. Thus, the LH design provides some information about the interior of the experimental region. Ye (1998) described a procedure to construct an Orthogonal Latin Hypercube (OLH) when its number of rows n is a power of 2 or a power of 2 plus 1, i.e. for $n = 2^m$ or $2^m + 1$, an OLH with $2m - 2$ columns can

be constructed. Later, Cioppa (2005) extended Ye’s procedure (Ye 1998) to incorporate more factors into the design matrix. They, however, noticed that the space-filling of these new designs is poor; therefore, they suggested sacrificing some of the orthogonality intentionally in order to achieve better space-filling while incorporating a greater number of factors. These designs are called Nearly Orthogonal Latin Hypercubes (NOLH). Sanchez (2005) has implemented an Excel program which produces an NOLH for up to 29 factors. Since NOLH designs are not readily available for more than 29 factors, a PB design augmented with a center point was used for exploring the experimental region for the CRC Model.

For 72 factors, PB requires 76 treatments (a combination of different levels of factors). This small experiment can be used to insure that the response function has equal variance over the experimental region. When the variances across treatments are not equal, the usual analysis of variance assumptions are not satisfied. Moreover, some of the factor screening methods designed for stochastic models assume homogeneous variance over the entire experimental region. If this assumption is not satisfied, a variance-stabilizing transformation is recommended and the experiment is run on the transformed data (Montgomery 2000).

We chose Cost, QALY, and Cost over QALY (Cost/QALY) as the responses to be screened. Our previous experience with the model revealed that the response variance is dramatically high. One approach to reduce the variation among the observations is to put observations into a batch and treat the batch mean as one observation. For each treatment, a total of 100 observations were obtained, grouped in 10 batches each containing 10 observations. Several statistical tests have also been proposed for testing the equality of variance. These tests can be viewed as formal tests of the hypotheses

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_a^2$$

$$H_1 : \text{above not true for at least one } \sigma_i^2$$

where a is the number of treatments. Montgomery (2000) recommends using “Bartlett’s test” when the normality assumption holds and “Modified Levene test” when the normality assumption is violated. Since Bartlett’s test is sensitive to the normality assumption, there may be situations where the Modified Levene test is more appropriate. Bartlett’s test is usually conducted at the significance level 0.01 or 0.001. Anderson et al. (1974) suggests rejecting the variance homogeneity assumption if the test is rejected at the significant level 0.001 and if the homogeneity test is accepted at $\alpha = 0.01$ level, no transformation is required. For all three responses, the p-values for both the Bartlett’s and Levene tests turned out to be less than 0.0001 and therefore we reject the variance homogeneity assumption.

3.3 Factor Screening Experiment

Among the factor screening methods discussed in section 2, CSB and its variants (CSB-X, CSFD, and Hybrid method) can operate when response variance is heterogeneous. However, due to the particular random sampling structure of CRC Model that limits sequential observations, none of those methods can be applied. 2^k Fractional Factorial designs assume that the response variance is homogeneous; however, with proper transformation we can stabilize the variance and obtain valid results.

The values for the “high” and “low” values of the variables used in the fractional factorial experiments for the factor screening were either upper and lower limits placed on the variables (most of these were Johnson S_B distributions) or upper and lower bounds generally ± 3 standard deviations. Negative lower bounds were truncated at zero.

A 2^k Fractional Factorial design of Resolution IV requires 256 treatments (design points) for screening 72 factors. This design can estimate the factors’ main effects clear of the second-order interaction effects. A total of 100 observations were obtained for each design point and grouped into 10 batches. The adjusted R^2 for the model fitted to Cost, QALY, and Cost/QALY are 0.8146, 0.7530, 0.8184, respectively. The p-values show that the generated models have fitted the corresponding data fairly good. To test the variance homogeneity assumption, we used the plot of actual response and residual by predicted response. Figure 1 shows the corresponding plots for Cost/QALY as response. As Figure 1 shows, the variance heterogeneity is obvious and corrective transformation techniques should be used. Cost and QALY showed the same behavior and required data transformation.

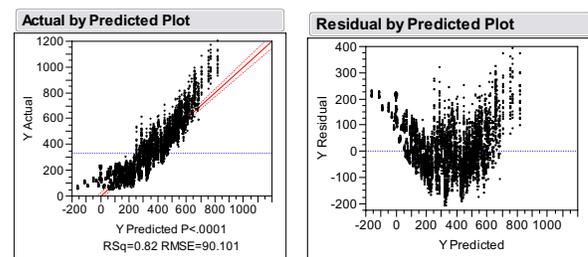


Figure 1: Actual by predicted and residual by predicted plots for Cost/QALY as response

We used Cox-Box method (Montgomery 2001) to find an appropriate transformation for the responses. Figure 2 shows the plot of actual and residual by predicted response for Cost/QALY after applying the Cox-Box transformation. Comparing Figure 1 and Figure 2 indicates that transformation has truly stabilized the response variance. Moreover, transforming data increased

the adjusted p-value from 0.8146 to 0.8504 for Cost, from 0.7530 to 0.7537 for QALY, and from 0.8184 to 0.8545 for Cost/QALY.

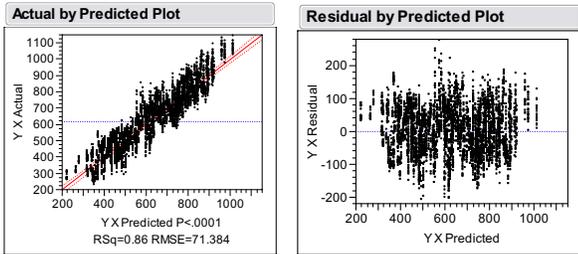


Figure 2: Actual by predicted and residual by predicted plots after transformation for Cost/QALY as response

Figure 3, Figure 4, and Figure 5 show the seven most important factors with highest absolute value of main effects for Cost, QALY, and Cost/ QALY, when Cox-Box transformation is applied to the data.

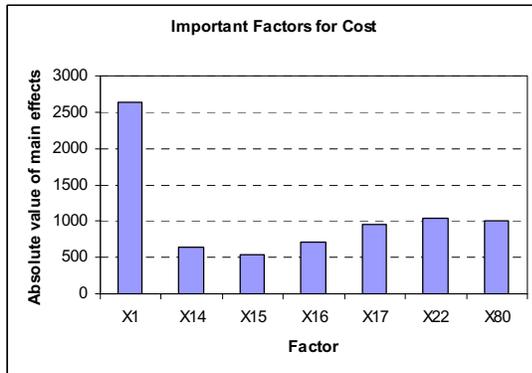


Figure 3: Important factors for Cost

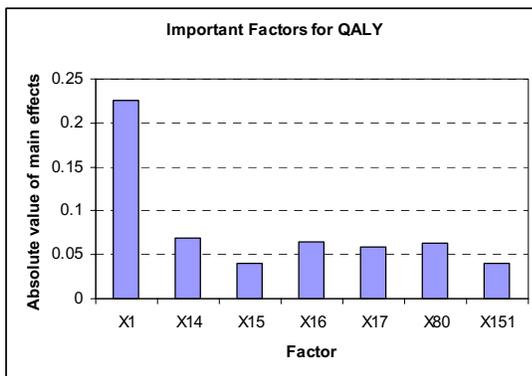


Figure 4: Important factors for QALY

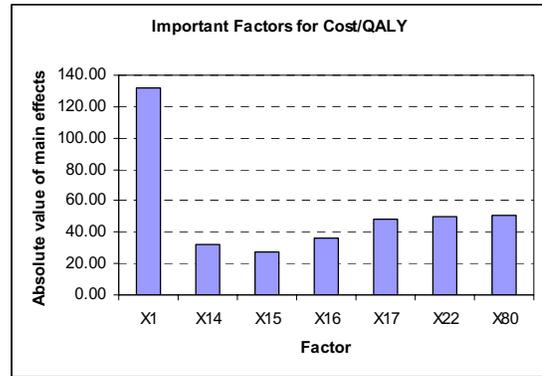


Figure 5: Important factors for Cost/QALY

Table 1 lists the name of the important factors with their corresponding screening IDs. Table 2 summarizes the important factors with their estimated main effects for Cost, QALY, and Cost/QALY. It should be pointed out that the estimated main effects are for the transformed metamodel. Nevertheless, since the Cox-Box transformation is a monotonic function, the signs of estimated main effects after transformation are the same as the signs of main effects in the original metamodel.

Table 1: The name of important factors

Screening ID	Factor Name
X1	RiskDistribution
X14	Time from Advanced adenoma to asymptomatic local CRC
X15	Time from Local to regional CRC
X16	Time from Regional to distant CRC
X17	Time from Asymptomatic to symptomatic CRC
X22	Probability of surgical (first month) mortality
X80	RiskAffects
X151	GeneralCompliance

Table 2: List of important factors with their estimated main effects

Screening ID	Cost	QALY	Cost/QALY
X1	2642.68	-0.226	132.12
X14	-644.45	0.069	-32.42
X15	-544.90	0.040	-27.76
X16	-702.61	0.065	-36.12
X17	949.55	-0.059	48.21
X22	-1041.54	UI	-50.15
X80	1011.43	-0.064	50.56
X151	UI	0.041	UI

UI: Unimportant effect

4 VERIFYING THE RESULT OF FACTOR SCREENING

To verify the results of factor screening for each method we performed the following steps:

1. A metamodel was generated for the factors detected as important by using a 2^k Fractional Factorial of Resolution V. In order to have more accurate estimates for the metamodel parameters, we applied “Rule of 2” (Wallace 1977), i.e. eliminating the effects with F-ratio less than 2 from the generated metamodel. The metamodels that are generated for only important factors will be called “restricted metamodels.” Also, in order to stabilize the response variance, we transformed the observed data according to the Cox-Box method.
2. The experimental region was sampled according to a Latin Hypercube Design (LHD) with 40 rows and 72 columns. To produce a LHD with good space-filling property, we generated 10000 random LH designs and chose the one with the highest minimum distance between the design rows. For each design point 100 observations were obtained and placed in batched of size 10. Then, to examine how well the generated metamodel can represent the response function over the entire experimental region, the response value corresponding to each treatment in the LHD was predicted by the restricted metamodel. Next, we used the *coefficient of determination* (Kleijnen et al. 2000) to quantify the goodness of fit of the metamodel.

The coefficient of determination is defined as

$$r^2 = 1 - \frac{\sum_{i=1}^n \sum_{r=1}^{m_i} (\hat{y}_i - y_{i,r})^2}{\sum_{i=1}^n \sum_{r=1}^{m_i} (y_{i,r} - \bar{y})^2}$$

where, \hat{y}_i is predicted value for treatment i , $y_{i,r}$ is actual response for r^{th} observation of treatment i , \bar{y} is mean of all observations, m_i is number of observations for treatment i , and n is the total number of observations for all treatments. r^2 equals one (perfect fit) if all n metamodel output equals their corresponding simulation outputs ($\hat{y}_i = y_{i,r}$, for all i and r). Because r^2 always increases as more regression variables are added (higher q), the Adjusted r^2 is introduced $r_{Adj}^2 = 1 - (1 - r^2)((N - 1)/(N - q))$ where, N is the total number of obtained runs and q is the number of regression variables. r_{Adj}^2 represent the amount of variation over the entire experimental region that has been accounted for by the generated metamodel.

For the generated metamodels, r_{Adj}^2 turned out to be 0.6561 for Cost, 0.6342 for QALY, and 0.6552 for Cost/QALY. Thus, the result of factor screening can be considered valid.

5 CONCLUSION AND FUTURE RESEARCH

Out of the 72 factors considered in the CRC model, only eight seemed to be consistently important in a factor analysis for cost, quality-adjusted life-years (QALY), and cost per QALYs. The fact that the same factors were important for both cost and QALY is surprising since each outcome is a very different measure of effectiveness. However, since both produce similar importance, it is then expected that cost per QALYs would have the same important factors.

Closer examination of the important factors reveals that the simulation is fundamentally driven by a few very important factors, although clearly many other factors contribute. These “driving factors” reveal some basic elements of the model that tend to be obscured by the many factors employed.

The risk distribution is seen to be especially important. The role of “risk” throughout the model is what determines not only the propensity to develop adenomas through incidence, but also influences the progression of adenomas to cancer. Risk is a highly transparent variable and has direct clinical relevance to physicians who must assess “risk” relative to individual patients. It is not surprising that risk is so strongly important.

The “time” distributions, namely time from advanced adenoma to asymptomatic local CRC, from local to regional CRC, from regional to distant CRC, and from asymptomatic to symptomatic all have similar importance with the later ones showing more impact. These factors are critical in the time stream from adenoma to cancer and whether these intermediate events occur before natural death. It is therefore easy to understand why these are important. However it is surprising that the time from invisible adenoma to visible adenoma and the time from visible adenoma to advanced adenoma is also not equally important.

The role of “RiskAffects” in the simulation model relates to the incidence and progression of the adenomas within individuals. Thus, like “risk” alone, these effects can greatly impact the activity of the disease and thus have important influence on the outcomes.

The importance of the probability of surgical mortality seems to impact cost and cost/QALY more than QALY simply because a surgical mortality has such a high cost. Likewise the variable “GeneralCompliance” appears to impact QALY while being relatively unimportant for cost and cost/QALY.

The overall value of this factor screening applied to the CRC simulation model provides insight into the “driving factors” in the model. Unfortunately the factors identified are factors associated with the natural history of disease and are not amenable to external control (such as screening age). In this case the factor screening amplified the critical role of certain basic variables, which were (for

the most part) recognized as fundamental. There are numerous other variables that have a clear importance but were not identified by factor screening based on the FF-IV.

The fact that many seemingly important variables do not manifest themselves in factor screening using fractional factorials suggests that there is some deficiency in the use of this approach for factor screening (at least for this case). We suspect that much of this inadequacy to identify other factors is related to the basic limitations of the fractional factorial from which importance is determined.

The fractional factorial based on 2^k determines its importance over the range from "low" to "high" values of the factors. Thus the range endpoints of these factors influence their chance of importance. Wan et al. (2004) show how the range endpoints can be determined if we can associate cost to each factor. However the factors in the CRC model do not lend themselves to this kind of estimation and the range endpoints had to be determined arbitrarily.

Also the CRC model appears to be highly non-linear with unknown higher-order interactions that compromise the FF-IV analysis. Recall that FF-IV can only estimate main effects clear of first-order interactions, however the simulation may produce important interaction effects that are not clearly estimated. Perhaps a means of analysis that can accommodate to these characteristics should be explored.

REFERENCES

- Anderson, Virgil L. and Robert A. McLean. 1974. *Design of experiments - a realistic approach.*: Marcel Dekker Inc.
- Bettonvil, Bert W. M. and Jack. P. C. Kleijnen. 1996. Searching for important factors in simulation models with many factors: Sequential bifurcation. *European Journal of Operational Research* 96: 180-194.
- Cheng, Russell. C. H. 1997. Searching For Important Factors: Sequential Bifurcation under Uncertainty. In *Proceedings of the 1997 Winter Simulation Conference*, 275-280.
- Cioppa, Thomas M. Experimental Designs for High-Dimensional Complex Models. 2005. United State Army Training and Doctrine Command Analysis Center.
- Johnson, Wesley O. and Joseph L. Gastwirth. 1999. Dual Group Screening. *Journal of Statistical Planning and Inference* 83: 449-473.
- Kleijnen, Jack. P. C. 1987. Review of Random and Group-Screening Designs. *Communications in statistics.Theory and methods* 16, (10): 2885-2837.
- Kleijnen, Jack. P. C. and Robert G. Sargent. 2000. A Methodology for Fitting and Validating Metamodels in Simulation. *European Journal of Operational Research* 120: 14-29.
- Li, C. H. 1962. A Sequential Method for Screening Experimental Variables. *Journal of the American Statistical Association* 57: 455-477.
- McKay, M. D., Beckman R.J., and W. J. Conover. 1979. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics* 21, (2): 239-245.
- Montgomery, Douglas. C. 2000. *Design and Analysis of Experiments*. 5 ed.: John Wiley & Sons.
- Montgomery, Douglas. C. 2001. *Design and Analysis of Experiments*. 5 ed.: John Wiley & Sons.
- Ottieno, J. A. M. and M. S. Patal. 1984. Two Stage Group-Screening Designs with Unequal A-Prior Probability. *Communications in statistics.Theory and methods* 13, (6): 761-779.
- Patal, M. S. and J. A. M. Ottieno. 1984. Two-Stage Group-Screening Designs with Equal Prior Probabilities and no Errors in Decision. *Communications in statistics.Theory and methods* 13, (9): 1147-1159.
- Patel, M. S. 1962. Group-screening with more than two stages. *Technometrics* 4: 209-217.
- Roberts, S. D., L. Wang, R. L. Klein, R. M. Ness, and R. S. Dittus. 2008. Development of a Simulation Model of Colorectal Cancer. *Transactions on Modeling and Computer Simulation*: to appear in January.
- Sanchez, Susan M. 2005. "NOLHdesigns spreadsheet." Available online via <http://diana.cs.nps.navy.mil/SeedLab/> [accessed June 5,2007].
- Sanchez, Susan M., Hong Wan, and Thomas W. Lucas. 2005. A Two-Phase Screening Procedure for Simulation Experiments. 223-230. *Proceedings of the 2005 Winter Simulation Conference*.
- Shen, Hua and Hong Wan. 2005. Controlled Sequential Factorial Design for simulation Factor Screening. 467-474.
- Shen, Hua and Hong Wan. 2006. A Hybrid Method for Simulation Factor Screening. 382-389.
- Tafazzoli, A., S. D. Roberts, R. M. Ness, and R. S. Dittus. 2005. A Comparison of Screening Methods for Colorectal Cancer Using Simulation Modeling. In *Proceedings of the 2005 Winter Simulation Conference*, ed. Kuhl, M. E., N. M Steiger, F. B. Armstrong, and J. A. Joines, 2236-2245. IEEE.
- Trocine, Linda and Linda C. Malone. 2000. Finding Important Independent Variables through Screening Designs: A Comparison of Methods. In *Proceedings of the 2000 Winter Simulation Conference*.

- Wallace, T. Dudley. 1977. Pretest Estimation in Regression: A Survey. *American Journal of Agricultural Economics* 59, (3): 431-443.
- Wan, Hong, Bruce E. Ankenman, and Barry L. Nelson. 2003. Controlled Sequential Bifurcation: A New Factor-Screening Method For Discrete-Event Simulation. In *Proceedings of the 2003 Winter Simulation Conference*, 565-573.
- Wan, Hong, Bruce E. Ankenman, and Barry L. Nelson. 2004. Simulation Factor Screening with Controlled Sequential Bifurcation in the Presence of Interactions, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208-3119, U.S.A.
- Wan, Hong, Bruce E. Ankenman, and Barry L. Nelson. 2005. Controlled Sequential Bifurcation: A New Factor-Screening Method for Discrete-Event Simulation, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208-3119, U.S.A.
- Ye, Kenny Q. 1998. Orthogonal Column Latin Hypercubes and Their Application in Computer Experiments. *Journal of American Statistical Association* 93, (444): 1430-1439.

AUTHOR BIOGRAPHES

REZA YAESOUBI is a PhD student in Industrial and Systems Engineering and Operations Research at North Carolina State University, where he received his Master of Science degree in December 2006. His research interests include conducting metamodeling analyses on the stochastic simulation models. His web page can be found via <http://www4.ncsu.edu/~ryaesou>.

STEPHEN ROBERTS is a professor in the Edward P. Fitts Department of Industrial and Systems Engineering at North Carolina State University. His research interests include simulation modeling methodology and the simulation modeling of medical decisions. His e-mail is roberts@ncsu.edu.