A MODEL FOR MAKING PROJECT FUNDING DECISIONS
AT THE NATIONAL CANCER INSTITUTE

NICHOLAS G. HALL
Ohio State University, Columbus, Ohio

JOHN C. HERSHEY
University of Pennsylvania, Philadelphia, Pennsylvania

LARRY G. KESSLER and R. CRAIG STOTTS
National Cancer Institute, Bethesda, Maryland

(Received August 1991; revision received January 1992; accepted March 1992)

This paper describes the development of a model for making project funding decisions at The National Cancer Institute (NCI). The American Stop Smoking Intervention Study (ASSIST) is a multiple-year, multiple-site demonstration project, aimed at reducing smoking prevalence. The initial request for ASSIST proposals was answered by about twice as many states as could be funded. Scientific peer review of the proposals was the primary criterion used for funding decisions. However, a modified Delphi process made explicit several criteria of secondary importance. A structured questionnaire identified the relative importance of these secondary criteria, some of which we incorporated into a composite preference function. We modeled the proposal funding decision as a zero-one program, and adjusted the preference function and available budget parametrically to generate many suitable outcomes. The actual funding decision, identified by our model, offers significant advantages over manually generated solutions found by experts at NCI.

The mortality rate from lung cancer in the United States increased 400% from 10 per 100,000 population in 1940, to 50 in 1990, in stark contrast to the falling or stabilized rates of all other major cancers (American Cancer Society 1990, p. 5). This year over 400,000 smokers will die prematurely, representing one out of every six deaths, and 87% of all lung cancer deaths, in this country (U.S. Department of Health and Human Services 1989, p. 161). Federal, state and local governments receive $9.6 billion annually from the sale of tobacco products (Tobacco Institute 1988, p. 4). However, the annual costs of tobacco use, including work absenteeism, lost productivity and increased health care costs, exceed $65 billion (Office of Technology Assessment 1985, p. 5).

At present, 27% of the adult U.S. population smokes, representing more than 50 million smokers. This epidemic is showing only slow signs of abatement, due primarily to the highly addictive nature of nicotine. It is not surprising, then, that the U.S. Surgeon General proclaimed cigarette smoking "the most important public health issue of our time" (U.S. Department of Health and Human Services 1982, p. xi).

In response to the mounting human and economic toll from tobacco smoking, the National Cancer Institute (NCI) has, since 1982, funded a series of studies and experiments designed to find effective mechanisms for reducing smoking prevalence (Greenwald and Sondik 1986). These have culminated in the planning of the American Stop Smoking Intervention Study (ASSIST), the largest public health initiative ever undertaken by the National Institutes of Health. The main goal of ASSIST is to put the nation on course toward the Healthy People 2000 goal of a one-half reduction in smoking prevalence rate.

Subject classifications: Decision analysis, applications: incorporating decision makers' preferences into optimization model. Health care: national anti-smoking campaign. Programming, integer, applications: optimization model to support funding decisions.

Area of review: OR PRACTICE.
from 30% of the population over 16 years of age in 1985, to 15% in the year 2000 (U.S. Department of Health and Human Services 1991, p. 140).

ASSIST will allocate approximately $114 million of federal money, and a substantial amount of private funds, into 15-20 states throughout the U.S. over a period of seven years. The purpose is to demonstrate that, with various combinations of policy changes, mass media information, and smoking cessation programs, smokers can be encouraged to stop smoking, and young people can be prevented from starting to smoke.

The process of issuing a complex request for proposals (RFP) and reviewing the resulting proposals defined an interesting decision problem: how to make awards among many competing contract proposals while balancing a number of considerations critical to the long-term viability and effectiveness of ASSIST. The final selection of awards was to be made by the director of the Division of Cancer Prevention and Control of NCI, with consultation from selected experts within the division. We refer to this group as the decision makers.

The size of the budget for ASSIST, and the large number of proposals submitted, generated a complex decision problem. For example, the number of possible ways of funding is far more than could be considered individually. In view of this, and the need to identify secondary criteria and allow them to influence the funding decision, decision makers at NCI decided that a formal modeling approach should be used. After the ASSIST RFP was issued, but well prior to the award date, one of the decision makers most involved in the project approached the third author to request assistance. The decision to formalize the site selection process was motivated by political issues and technical complexity. Because of the size of the procurement, and the prominent position of NCI in the public health community, some decision makers felt that significant political pressure might be placed on NCI during the selection process. A structured, formalized decision process was one method to avoid such pressure. Furthermore, there were three considerations in making the final awards: technical merit, cost, and secondary criteria mentioned in the RFP. Balancing these considerations would have been very difficult for even the most experienced of the decision makers. This paper describes the methods we used to help the decision makers at NCI select the sites for the ASSIST project.

Formalizing the criteria in an RFP presents an interesting challenge when an optimization model is to be used. We use a modified Delphi approach, including gathering an informal group of decision makers to identify key criteria to investigate, a mail questionnaire to elicit quantitative impressions of the relative importance of criteria, structured in-person interviewing to develop quantitative estimates of preference, and regression analysis to produce preference estimates. This interactive approach gives the decision makers a feeling of involvement, as well as a chance to see different versions of the criteria over time.

Our role was to provide the decision makers with a short list of combinations of states that would maximize the budget's ability to fund the states with the most highly evaluated proposals, while also meeting other criteria mentioned in the RFP. The decision makers were then to assess the various options, consider the tradeoffs, and make the decision.

In Section 1, we describe the background and application process for ASSIST. In Section 2, we describe how we formalize the intentions of decision makers with respect to primary and secondary funding criteria for use in a mathematical model described in Section 3. Our results, and the methods we develop for their presentation to decision makers, are described in Section 4. Finally, Section 5 summarizes the advantages and limitations of our approach with respect to proposal funding decisions, and assesses the reactions of the decision makers at NCI to our study.

1. BACKGROUND

The strategy behind the ASSIST project is to combine federal and local funding, in order to encourage a sense of ownership of the project at each selected location. This, in turn, should forge alliances that will last well beyond the seven year funding period. Before a proposal could be considered eligible for an award, certain mandatory criteria had to be met.

Proposers had to be from one of the 50 states or the District of Columbia, or one of a selected list of large metropolitan areas, and had to be working actively with a voluntary health agency such as the American Cancer Society.

Each proposal was subjected to a scientific peer review, using the following technical criteria: management approach, organizational experience, experience and capabilities of coalition member groups, experience and capabilities of key personnel, and availability of facilities and equipment.

More than 35 states (and no metropolitan areas) initially responded to the RFP by submitting a proposal. Three review meetings were held, the final one taking place after submission of answers to detailed
questions of clarification from the review committee. By the conclusion of the third meeting, the final list of eligible states with technically acceptable proposals was reduced to 23. Technical scores at the end of this process ranged from a high of 958 (out of 1,000) to a low of 598. Technical merit was a primary consideration in giving awards. However, for a variety of reasons discussed in Section 2, neither the technical scores, nor the ranking of the proposals derived from these scores, was an appropriate measure for use in the model. The design of an alternative measure with several desirable properties is described in Section 2.

As with any RFP, the federal government retained some discretion about what services should be provided and by whom. In the ASSIST RFP, the services and the provider qualifications were clearly stated, but specific language to allow flexibility was also inserted. This language stated that the federal government reserved the right to make its selections not only on the basis of technical scores, but also using considerations of “smoking prevalence and geographical distribution in order to obtain approximately equal geographical distribution throughout the United States.”

These last two factors defined much of our task, namely to determine what those factors meant to the decision makers at NCI, and to formalize them into a mathematical model for generating a variety of possible solutions. However, for contractual reasons, and because the technical criteria represent the factors likely to lead to a successful contract, the technical evaluation of the proposals had to remain the primary criterion in determining which states would receive awards.

A possible danger of introducing flexibility into the decision making process is that the process can become politicized. To protect our own objectivity, we used numerical codes to identify the states throughout our modeling work and even in the presentations we made to the decision makers.

We are aware of only two related studies for making proposal funding decisions. Sarin, Sicherman and Nair (1978) develop a multiattribute utility approach, without the use of an optimization model, and describe an application to a solar energy experiment. Peerenboom, Buehring and Joseph (1989) study a decision problem for a synthetic fuels facility, in which a hierarchy of decisions is required. A dynamic programming algorithm is used to make budget allocations between the different levels. In our model, by contrast, we use the formal model to make the detailed selection of proposals. Moreover, we implement secondary criteria directly into the optimization model as constraints. Another major difference is that the exact final decision implemented was identified by our formal analysis.

2. FORMALIZING THE CRITERIA

There were two key issues in formalizing the intent of the RFP. The first relates to the ranking of the proposals by the cardinal measure of technical score, or by raw rank (the ordinal measure derived from technical score). The other involves formalizing the secondary criteria mentioned in the RFP. With regard to ranking the proposals, and as a result of several discussions with the ASSIST project management team and staff from the contracts office of NCI, we concluded that both the above measures possess undesirable properties. The cardinal measure of technical score is considered less important than the rank of the proposals. Differences in technical score, beyond those expressed in the rank itself, are of relatively little importance. On the other hand, for contractual reasons, funding out of rank order would need to be justified by reference to secondary criteria.

The raw rank of the proposals, however, was also an inadequate measure, because funding in rank order is considered much more important among the top-ranked proposals than among those ranked lower. For example, other characteristics being equal, the loss from funding the second-ranked proposal in place of the first-ranked is much greater than similarly for the tenth- and ninth-ranked proposals, respectively. Another problem with raw rank is that an increase or decrease in all ranks will change the ratios of two ranks. Since it is the ratios of the coefficients of a linear optimization model that determine the solution, an objective function that allows such changes in ratios is not suitable.

We therefore designed our own monotonic rank function measure, \( V(\cdot) \), with the desired properties. Those properties, assuming that the proposals are ranked from \( n \) (best) to \( 1 \) (worst), where \( n \) is the number of proposals, are:

1. **Strict convexity:** differences in rank are more significant among the high rank values, or \( V(i + 1) - V(i) > V(j + 1) - V(j) \) for \( n > i > j \geq 1 \).
2. **Constant tradeoff rate:** if the value of an increase in rank from \( i \) to \( i + a \) is equal to that of an increase from \( i \) to \( i + b \), then this should be true for all \( b + 1 \leq a \leq n \). This property is analogous to constant risk aversion for utility functions (Keeney and Raiffa 1976, pp. 165–168).
3. Invariance to addition or subtraction: the relative preference (ratio) between one group of proposals and another should not change if the rank of all proposals increases or decreases by the same constant.

These three properties imply that the rank function should take the form:

\[ V(y) = \frac{\exp(cy)}{\exp(c)}, \]

where \( y \) is the rank of the proposal, and \( c \) is a non-negative parameter to be estimated. Note that the denominator normalizes this function to yield \( V(1) = 1 \).

To estimate \( c \), we asked one of the decision makers most involved in the project questions such as the following: "We want to find out something about the disutility of funding proposal number 15 rather than number 18. What would be the ranking, below 15, for which moving there from 15 would be equivalent to moving from 18 to 15?" The answer to this question (which was 11) implies a value of \( c = 0.083 \), which can be found by successive approximation. A total of seven questions of this type, using both increasing and decreasing ranks, yielded a mean \( c \) value of 0.094. The estimated standard error of the mean was 0.022. Using the mean value, our rank function is \( V(y) = \frac{\exp(0.094y)}{\exp(0.094)} \). The rank function scores thus calculated appear in Table I. Only one decision maker was asked for help, because of a lack of time in building the model and a lack of technical knowledge about this area among the decision makers. In Section 4, we demonstrate that the results are not sensitive to the estimation of \( c \). As we discuss below, these rank function scores constitute the primary basis for making funding decisions.

The other issue in formalizing the intent of the RFP involves the secondary criteria mentioned therein. Our purpose here was to identify a small number of criteria, and their relative importance, for inclusion in a composite preference function. A questionnaire was designed to elicit such information from the decision makers. To determine what "geographical distribution" and "smoking prevalence" meant to NCI, we met with two key decision makers, gave them a preliminary list of our proposed questionnaire items, and requested their ideas on what should be asked of a larger group of decision makers. From these meetings, we determined that geographical distribution was an unambiguous concept needing no further specification, and that setting a minimum requirement of at least three states from each of the Census Bureau's four regions would be an appropriate standard.

However, smoking prevalence had numerous interpretations and proved much more difficult to specify. Smoking rates vary in this country according to socioeconomic status, race and age, and also vary over time. The main question was whether decision makers should select states with the highest overall smoking prevalence rates (or states with large numbers of smokers), or alternatively, states that would reflect a distribution of rates by demographic and historical factors. The latter option was more appealing from a scientific point of view in that the results would be more generalizable. However, a staff person in the ASSIST program might prefer the first option in that more smokers would be reached. We decided to offer both these options on three items in the questionnaire. Those items deal with the rate of decline of smoking prevalence between 1985 and 1989, the per capita consumption of cigarettes, and the cost of the proposal per smoker reached.

The decision makers were then contacted by mail. We explained what we were proposing to do and why their help was needed. Six of the seven individuals agreed to participate, and were sent a questionnaire (Table II) that employed a three-point, forced-choice scale. Any item with a mean score greater than or equal to two would be deleted.

The questionnaire was developed exclusively for this project. Face validity was established through consensus of a panel of NCI experts in related areas. No
Table II
Delphi Round 1

Please circle the number that most closely corresponds to your estimate of how important the criterion is in choosing ASSIST award sites in addition to the technical score.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Very important</th>
<th>Somewhat important</th>
<th>Not important</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>A site should be chosen if it:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Has a high proportion of Blacks</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1.8 0.374</td>
</tr>
<tr>
<td>2. Has a high proportion of Hispanics</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1.8 0.374</td>
</tr>
<tr>
<td>3. Has a high proportion of Native Americans</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2.4 0.245</td>
</tr>
<tr>
<td>4a. Has experienced a steep decline in smoking prevalence in 1985–1989</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2.75 0.25</td>
</tr>
<tr>
<td>OR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4b. Provides a distribution (&quot;a spread&quot;) of changes in smoking prevalence rates in 1985–1989</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1.5 0.5</td>
</tr>
<tr>
<td>5. Reaches a large number of smokers</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1.0 0</td>
</tr>
<tr>
<td>6a. Has high per capita consumption of cigarettes OR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6b. Provides a distribution (&quot;a spread&quot;) of per capita consumption of cigarettes</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2.25 0.479</td>
</tr>
<tr>
<td>7a. Has the lowest cost per smoker reached OR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7b. Provides a distribution (&quot;a spread&quot;) of expenditure per smoker</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1.5 0.289</td>
</tr>
<tr>
<td>8. Has a high quit rate</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3 0</td>
</tr>
</tbody>
</table>

COMMENTS: Please write your comments on the reverse side and number them according to the criterion.

effort to establish reliability was made due to the narrow focus of the questions. The panel of experts agreed that the way the questionnaire was worded would be completely clear to the decision makers for whom it was designed. They also agreed that, in order for a criterion to survive the first Delphi round, it would need fairly strong consensus, i.e., a rating of at most 2.0, or "somewhat important."

The mean and standard error for each question are also shown in Table II. Having a high proportion of Native Americans in the sample was not considered a high priority. This was not because of a lack of concern about high smoking rates among Native Americans, but because finding sufficient numbers of subjects for statistical analysis in a national survey is very difficult. The items on smoking rate decline and cost per smoker both generated a high preference for a distribution of rates. The item on per capita consumption of cigarettes found a higher preference for a high rate of consumption than for a distribution of rates.

The next round examined the items that remained from the first round, but in a different manner. The decision makers were asked to assume that all funding decisions except one had been made for each question. Each question pitted two hypothetical qualifying states against one another, and respondents were asked to indicate their preferences between the two states in each pair, on a 5-point scale. This stage of the survey was conducted through face-to-face interviews to allow for maximum understanding of the choices. Interviewees were encouraged to ask questions of the interviewer, and were allowed to change answers at any time.

The questions were designed to produce a simple, additive, multiattribute preference rating for each candidate state. In one section of the questionnaire, for example, the pairs of states in each question presented different numbers of smokers and nonsmokers in the states to be compared. We performed a least-squares analysis, using the answers to these questions as the dependent variable, and assuming that the 5-point scale was an interval preference scale ranging from −2 to +2. The intercept was forced to have a value of zero on the converted scale, signifying no preference
when there were no differences in the independent variables. The best fit was given by the function:

\[
\text{Preference} = 2.07S - 0.45N,
\]

where \( S \) is the number of smokers (in millions) and \( N \) is the number of nonsmokers (in millions). This function favors states with high smoking rates and, to some degree, states with large populations, since all states have a smoking rate greater than \( 0.45/(2.07 + 0.45) = 0.179 \). This is consistent with the "qualitative" interpretation of the first round results. The answers to other paired comparisons in the second round gave us the information we needed to complete the constraint set for the model, as:

1. Decision makers wanted a distribution of smoking rates represented in the final set of chosen states, and they were indifferent between adding a state that was a "high" outlier versus a state that was a "low" outlier, apart from their general preference for higher smoking rates. Accordingly, we added a constraint ensuring that at least two states would be chosen from each quartile of the distribution of smoking rates across the 50 states and the District of Columbia. We chose two states because that number was apparently feasible, and a better representation than only one.

2. Decision makers wanted a distribution of decline in smoking rates for the past few years, and they were indifferent between adding a state that was a high outlier versus a state that was a low outlier. Therefore, we again added a constraint ensuring that at least two states would be chosen from each quartile of the distribution across the 50 states and the District of Columbia. The reasons for the choice of two states were similar to those above.

3. The proportions of blacks and Hispanics in state populations were of more importance than the absolute numbers.

4. If the proportion of blacks and Hispanics were to be included in the preference function, the best fitting function would be:

\[
\text{Preference} = 2.07S - 0.45N + 12.3M,
\]

where \( M \) is the proportion of the state's population that is black or Hispanic.

For the analysis that follows we did not explicitly include this final term that incorporates the preference for high rates of blacks and Hispanics in the definition of the preference score. This is because, although the decision makers consider this criterion important, contractual reasons prevent it being included formally in our model. The use of this criterion in our recommendations is discussed in detail in Section 4.

3. MATHEMATICAL MODEL

The overall purpose in formulating and solving a model for the project funding problem is to provide a short list of solutions. Those solutions should be within, or very close to, the estimated budget figure, and also among the best available solutions with respect to total rank function score and other important criteria. Decision makers at NCI will then have the option to select from among the recommended solutions, or seek additional information from the model.

One advantage of using an optimization model is that with 23 proposals the number of combinations that could be funded and that would satisfy the constraints is both large and difficult to identify. Optimization models allow quick solution of such problems. The use of heuristic approaches frequently means giving up some value with respect to the identified criteria, through the selection of a locally, but not globally, optimal solution. This may lead to inefficient use of available funds. A simulation procedure might also identify optimal solutions, but in a less direct and more time-consuming way. Since, as we shall discuss, the uncertainties in the problem require the solution of many variations of the same model, such an approach would quickly become computationally unworkable.

The ASSIST project was funded as a procurement or a contract for specified goods and services. A solution involving partial funding would not have been permitted by the NCI contracts procedure. Thus, the choice of a linear zero-one optimization model was identified as appropriate. The assumptions required by such models are well documented in, for example, Hillier and Lieberman (1980). The most critical assumption here is additivity. That is, the objective function and constraints need to represent the problem through the use of linear additive terms.

In many linear optimization models, the relative attractiveness of two opportunities is unchanged by the arrival or departure of a third. Our purpose in designing \( V(\cdot) \) in Section 2, however, was to define a value, not so much for a proposal itself, but for its ranking among all the proposals. The relative ranking of some other proposals would change if a proposal that is not ranked first or last arrives or departs. Consider the values of \( V(i) \) and \( V(j) \) for any \( i > j \) in Table I. If a new proposal with rank \( k, i > k > j \)
arrives, $V(i)$ is increased and $V(j)$ is unchanged. Alternatively, if $k > i$ or $k < j$, the relative score $V(i)/V(j)$ is unchanged from the third property (invariance) of $V(\cdot)$. Note that in a linear optimization model it is the relative coefficients of variables that determine the solution. The first two properties of $V(\cdot)$, convexity and constant tradeoff rate, are designed to generate relative "point" values that can be considered additive, and that capture the preference for funding highly ranked proposals. The constraints are mainly of a type for which additivity is a very natural assumption. The only exception is the preference constraint, for which the preference function coefficients had to be computed in such a way that they could be considered additive. The discussion in Section 2 shows that each preference coefficient is additive because each is a linear combination of two additive quantities.

It was made clear throughout the development of the project that the rank function score of proposals was to be given primary importance in funding decisions. We therefore chose as our objective function the maximization of rank function score over the projects funded. Other important criteria such as budget, diversity in smoking prevalence, and diversity in decline in smoking rate, and diversity in geographical area, were to be included as constraints. Some parameters of the model were not well specified. For example, the initial budget estimate of $114$ million was subject to the uncertainty that accompanies any multiple-year funding commitment in the government, where budgetary changes occur year by year. Moreover, it would be useful to have a range of possible solutions at different budget levels, so that if an opportunity existed for significant improvement in the solution, a case could be made for an increase in budget. The budget was therefore represented by a "soft constraint," which we modeled by allowing the right-hand side value to vary parametrically.

Another soft constraint represents the preference function. The purpose of the preference function constraint in our model is to identify the tradeoffs between total rank function score and the other criteria considered important. We avoid incorporating the rank function score with the other criteria because it is given a higher priority in funding decisions. As the amount of "preference points" required by the constraint increases, the greater the compromise of total rank function score that occurs. Highly ranked proposals are being replaced by those ranked lower, but with a higher preference value. If we fix the available budget, we can construct a tradeoff graph between various levels of preference function requirement and maximum achievable total rank function score. This graph typically has some points at which an increase of given size in the preference function requirement is associated with a smaller decrease in total rank function score than elsewhere. Such points are promising candidates for our short list of recommended solutions. This parametric analysis for the preference function requirement, with a fixed available budget, has been an invaluable tool for identifying a variety of interesting solutions. At some budget levels, as many as nine different solutions were identified by this method. The use of different values for the available budget similarly identifies additional candidates for the short list.

A formal statement of the exact model used appears in the Appendix. We now describe an example formulation. This example has all the features of the real model. However, the number of proposals, eight, is fewer. The numbers used in the example in Table III and the right-hand side values in the constraints, are for illustrative purposes, and are not taken from the real model.

The constraint requirements in this example are: minimum total preference points among funded proposals: 17.0; maximum available budget: $37.0$ million; minimum number of funded proposals in each region: 1; minimum number of funded proposals in each smoking prevalence quartile: 1; minimum number of funded proposals in each decline in smoking rate quartile: 1. The two values that will be changed

<table>
<thead>
<tr>
<th>Table III</th>
<th>Numbers Used in the Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item</td>
<td>Number</td>
</tr>
<tr>
<td>Proposal No.</td>
<td>1  2  3  4  5  6  7  8</td>
</tr>
<tr>
<td>Rank Function</td>
<td>1.9 1.8 1.6 1.5 1.3 1.2 1.1 1.0</td>
</tr>
<tr>
<td>Preference Coefficient</td>
<td>6.3 4.5 8.0 5.2 4.7 7.0 9.1 7.7</td>
</tr>
<tr>
<td>Budget ($m)</td>
<td>9.8 7.4 4.9 3.9 8.1 6.1 7.3 5.6</td>
</tr>
<tr>
<td>Region</td>
<td>1  3  1  2  4  4  3  2</td>
</tr>
<tr>
<td>Smoking Quartile</td>
<td>3  1  1  2  3  3  4  2</td>
</tr>
<tr>
<td>Decline Quartile</td>
<td>1  3  4  4  2  1  4  3</td>
</tr>
</tbody>
</table>
parametrically are the minimum number of preference points required (presently 17.0) and the maximum available budget (presently $37.0 million).

Let $X_j = 1$ if proposal $j$ is funded, $X_j = 0$ otherwise, $j = 1, \ldots, 8$. For the values shown, the formulation that we solve is:

**Maximize**

$$1.9X_1 + 1.8X_2 + 1.6X_3 + 1.5X_4 + 1.3X_5 + \cdots + 1.2X_6 + 1.1X_7 + 1.0X_8$$

subject to

$$6.3X_1 + 4.5X_2 + 8.0X_3 + 5.2X_4 + 4.7X_5 + \cdots + X_1 + X_3 + \cdots$$
$$X_2 + X_3 + \cdots + X_4 + \cdots$$
$$X_1 + \cdots + X_5 + \cdots$$
$$X_2 + \cdots + X_6 + \cdots$$

(1)

$$7.0X_6 + 9.1X_7 + 7.7X_8 \geq 17.0$$
$$6.1X_6 + 7.3X_7 + 5.6X_8 \leq 37.0$$

(2)

$$X_6 \geq 1$$
$$X_7 \geq 1$$
$$X_8 \geq 1$$
$$X_6 \geq 1$$
$$X_7 \geq 1$$
$$X_8 \geq 1$$

$$X_1, \ldots, X_8 = 0 \text{ or } 1.$$ 

In this formulation, constraint 1 ensures that the proposals funded must achieve at least the minimum required total preference score of 17.0. Constraint 2 requires that the budget estimate of $37.0 million must not be exceeded. The remaining twelve constraints in the example divide into three groups of four constraints each. The first group requires the funding of at least one proposal from each of the four major geographical regions of the country. The second group requires the funding of at least one proposal from each smoking prevalence quartile. The third group similarly requires representation from each quartile for decline in smoking rate.

This model was solved on an IBM 3090 computer, running under the VMS/XA operating system, and using the commercially available software package LINDO version 5.0. A FORTRAN subroutine interacts with the software to solve the problem repeatedly for a variety of right-hand side values in (1) and (2). The real version of the model, which is shown in the Appendix, has 23 variables and 14 constraints, and the solution of a single instance requires about half a second of elapsed time. This time includes specification of the next formulation to be solved, setting it up in LINDO format, and the output of several pieces of information from the solution. The format and interpretation of the results obtained, as well as some sensitivity analyses that we performed, are described in Section 4.

### 4. RESULTS

Our computer model generated a large number of different solutions, which at first seemed surprising, because the estimated budget was about 77% of the budget that would have been needed to fund all 23 remaining proposals. Initially, the optimization model was run for large budget and preference score increments, before we focused on more detailed analysis. The best estimate of available budget was $114 million (versus $37.0 million in the example), and solutions were generated for budgets from $110 million through $118 million in increments of $0.5 million. A preference score range from 15 to 19 points (versus 17.0 in the example) was used, with increments of 0.1 points. In a series of runs, this generated a variety of solutions for any budget value. For each budget value, a graph was drawn to represent the solution sets. In Figure 1, each point corresponds to the funding of a combination of states, and its location on the graph shows its total rank function score and preference score. Lines are used to connect the combinations which were

![Figure 1](image-url)  
**Figure 1.** ASSIST site selection solution sets.
Table IV

ASSIST Solution Sets

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Straight</th>
<th>Straight</th>
<th>(A)</th>
<th>(B)</th>
<th>(A)</th>
<th>(B)</th>
<th>(A)</th>
<th>(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td># States</td>
<td>17</td>
<td>18</td>
<td>114-1</td>
<td>114-2</td>
<td>114-3</td>
<td>114-4</td>
<td>114-5</td>
<td>114-6</td>
</tr>
<tr>
<td>Total Population*</td>
<td>83.51</td>
<td>86.75</td>
<td>83.19</td>
<td>83.78</td>
<td>85.90</td>
<td>87.43</td>
<td>87.10</td>
<td>89.62</td>
</tr>
<tr>
<td>Total smokers*</td>
<td>18.15</td>
<td>18.88</td>
<td>18.05</td>
<td>18.16</td>
<td>18.68</td>
<td>19.04</td>
<td>19.09</td>
<td>19.65</td>
</tr>
<tr>
<td># Hispanics</td>
<td>3.61</td>
<td>3.79</td>
<td>4.17</td>
<td>4.18</td>
<td>4.35</td>
<td>3.83</td>
<td>3.76</td>
<td>3.85</td>
</tr>
<tr>
<td># Black, Quart-1b</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td># HISP, Quart-1b</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>% Blacks</td>
<td>12.1</td>
<td>14.4</td>
<td>12.09</td>
<td>11.90</td>
<td>10.75</td>
<td>11.66</td>
<td>11.91</td>
<td>11.77</td>
</tr>
<tr>
<td>% Hispanics</td>
<td>4.3</td>
<td>4.4</td>
<td>5.02</td>
<td>4.99</td>
<td>5.06</td>
<td>4.38</td>
<td>4.31</td>
<td>4.29</td>
</tr>
<tr>
<td># from region</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Total Points</td>
<td>70.32</td>
<td>71.92</td>
<td>70.66</td>
<td>70.42</td>
<td>70.32</td>
<td>69.16</td>
<td>68.29</td>
<td>67.85</td>
</tr>
<tr>
<td>Preference Score</td>
<td>16.45</td>
<td>17.12</td>
<td>16.35</td>
<td>16.49</td>
<td>17.02</td>
<td>17.32</td>
<td>17.66</td>
<td>18.13</td>
</tr>
<tr>
<td>Budget*</td>
<td>108.87</td>
<td>116.30</td>
<td>112.447</td>
<td>113.890</td>
<td>113.476</td>
<td>113.780</td>
<td>113.277</td>
<td>113.867</td>
</tr>
</tbody>
</table>

State by Rank

| Rank | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. | 11. | 12. | 13. | 14. | 15. | 16. | 17. | 18. | 19. | 20. | 21. | 22. | 23. |
|------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
|      | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |

* In millions.

b Number of states which are in the top quartile of minority population proportion.

c Millions of dollars.
Table IV—Continued

<table>
<thead>
<tr>
<th>114-7</th>
<th>114-8</th>
<th>114-9</th>
<th>115.5-1</th>
<th>115.5-6</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>16</td>
<td>16</td>
<td>18</td>
<td>17</td>
</tr>
<tr>
<td>90.18</td>
<td>89.94</td>
<td>90.40</td>
<td>86.84</td>
<td>90.12</td>
</tr>
<tr>
<td>19.82</td>
<td>19.86</td>
<td>19.94</td>
<td>18.87</td>
<td>19.78</td>
</tr>
<tr>
<td>10.58</td>
<td>11.97</td>
<td>12.83</td>
<td>10.19</td>
<td>10.57</td>
</tr>
<tr>
<td>3.88</td>
<td>4.03</td>
<td>4.00</td>
<td>4.31</td>
<td>4.36</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>11.73</td>
<td>13.31</td>
<td>14.19</td>
<td>11.73</td>
<td>11.73</td>
</tr>
<tr>
<td>4.30</td>
<td>4.48</td>
<td>4.43</td>
<td>4.96</td>
<td>4.83</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>62.09</td>
<td>60.30</td>
<td>50.99</td>
<td>70.99</td>
<td>67.95</td>
</tr>
<tr>
<td>18.33</td>
<td>18.48</td>
<td>18.52</td>
<td>17.13</td>
<td>18.28</td>
</tr>
<tr>
<td>111.692</td>
<td>113.653</td>
<td>113.632</td>
<td>115.146</td>
<td>114.485</td>
</tr>
</tbody>
</table>

identified using the same budget level, but different requirements for preference score, in the formulation. Clearly, points that are higher and farther to the right are more desirable. The lines for the $112.5, $114, and $115.5 million budgets are shown in Figure 1.

Several solutions are tabulated in Table IV. They are shown in the format presented to decision makers. The first row of Table IV indicates how each solution was found. For example, solutions 114-1 through 114-9 are nine solutions found at a budget of $114 million by using different preference function requirements. For simplicity, we display only two of the nine solutions found at a budget of $115.5 million. The top part of the table shows key information about each solution, in particular the total number of smokers reached, the total population reached, preference score, and total rank function score. The checkmarks in the second part of the table identify the states to be funded. Note that the states are ordered by decreasing rank function score. This provides useful information to the decision makers, because skipping over a state in order to fund one ranked lower has to be justified on the basis of information in the RFP.

Table IV also shows what happens if funding decisions are made exclusively in rank function score order. The first two columns show how poor these solutions are. Funding the first 17 states uses only $108.9 million of the budget, and does not satisfy the criterion of including at least three states in each geographical region. In fact, it does not even provide two states per region. Alternatively, funding the first 18 states exceeds the budget by $2.3 million, and still does not satisfy the geographical constraint. The solution labeled 115.5-1, in which the first 16 states and numbers 20 and 22 are funded, was suggested by an expert decision maker at NCI. This solution satisfies the geographical constraint, but exceeds the budget by more than $1.1 million and does not have a high preference score. Several other solutions reach more smokers and more total population, and do so at lower cost.

We decided to present three appealing solutions (corresponding to points A, B and C in Figure 1) to the decision makers, while informing them that there were many more solutions available for examination. Point A is an extreme point solution that corresponds to maximizing total rank function points for a budget of $114 million, with no regard for preference score. Note from Figure 1 that both points B and C have a total rank function score that is higher than we might expect at their respective preference score levels, in line with the discussion in Section 3. By funding states 19 and 20 at point B instead of states 13, 16, and 22.
at point \( A \), the model reaches an additional 1.6 million smokers, still within the budget of $114 million. However, it was point \( C \), described in the last column of Table IV, which had the most appeal because of its coverage of still more smokers and its geographical distribution. Another attractive feature was that four states funded at point \( C \) are in the top quartile for proportion of black population, and similarly for Hispanic population. A case was made to justify the relatively small increase of less than $0.5 million and less than 0.5% in budget, in order to achieve this very appealing solution. Thus, the final decision was made in favor of the solution represented by point \( C \).

As can be seen from Table IV, the solution that the model helped to identify at point \( C \) enables the inclusion of an extra 3.28 million people, including 910,000 smokers, in the ASSIST project. It also saves $661,000, compared to the expert’s solution. The U.S. Government estimates that ASSIST will prevent 1.2 million smoking-related deaths over seven years (New York Times 1991). Given that estimate, an approximate contribution of our modeling effort within that overall figure is at least 43,500 lives.

We also made use of sensitivity analyses of potentially critical assumptions. Such analyses can be performed quickly. For example, we examined how the solutions were affected by increasing or decreasing the value of \( c \) in the rank function by one standard error (0.022). An increase in \( c \) has no effect on the solutions. A decrease in \( c \), however, yielded one new solution and eliminated another, but neither of these was considered a competitive candidate.

We performed sensitivity analyses on the specification of the rank function, and on the three sets of constraints that specified minimum representation in selected quartiles. As we describe in Stotts et al. (1992), the results were quite robust except with respect to the constraints on geographic distribution. When only two states, or as many as four states, were to be represented from each region of the country, the ability to achieve the other objectives of the program was markedly affected. Interestingly, however, the formulation requiring only two states per region failed to provide a solution that dominated point \( C \) in Figure 1 on the dimensions of rank function score and preference score.

5. DISCUSSION AND CONCLUSIONS

Since many proposal funding decisions need to be made annually, we anticipate that there will be numerous opportunities to make use of our methodology. We see several advantages of optimization models in this context. First, the abstraction of the model provides anonymity for the proposals, which limits the ways in which political and personal influence may enter into the decision making process. Second, even with only 23 active proposals, and more particularly with a larger number, the optimization model can identify combinations of options which even the most expertly guided manual process would frequently miss. Third, the ease of computer solution for simple optimization models permits parametric analyses which are ideal for investigating tradeoffs at different budget levels, and using different interpretations of the (not always precisely stated) conditions in a request for proposals.

That the decision makers chose one of the solutions proffered by our team was prima facie evidence of a successful effort. But, more important, the decision makers spent time interacting with the data supplied by the model, and contrasted it with the NCI experts’ recommendations. The model’s solution required the decision makers to skip over certain states and fund others, in order to provide better geographic distribution and smoking prevalence at the expense (although minimal) of rank. Although this had been NCI’s intention at the outset, it was the formal structure of the decision process that allowed the decision makers to feel comfortable doing so. In making the final decision, the decision makers showed a clear understanding of the model and its process. We noticed no explicit reservations on the part of the decision makers, and are convinced similar efforts will be supported in the future.

While the ASSIST budget of about $16 million per year is large, it is slightly less than 1% of NCI’s current overall annual budget. Since this is a relatively small percentage, it is possible to plan the necessary expenditures over several years, even in times of fiscal uncertainty. However, the U.S. Government reserves the right to cancel a contract at any time if funds are not available.

If the optimization model were provided to potential applicants preparing proposals, they could, in general, shape those to NCI’s needs. Although this could offer some advantages, there are also potential disadvantages. The most notable is the possibility that some states suitable for funding might not apply, because they might feel unable to compete on an even basis with others. In this situation, smaller states might feel this. It is clearly in the interests of NCI that as many states as possible should apply.

The reactions of decision makers to our analysis were very favorable. In particular, the tradeoff curves (Figure 1), and the presentation of a variety of detailed
solutions in a compact format (Table IV), proved to be valuable decision making tools. The fact that we had examined the sensitivity of our results with respect to several of the parameters used was important in giving decision makers confidence to choose from among the results we generated. This approach is currently being considered at The National Institutes of Health for making awards in an even larger study.

There were potential pitfalls to our modeling approach, however. For example, decision makers sometimes provided inconsistent responses that could be difficult to reconcile. A variety of different approaches may be needed in similar situations. Careful investigation of what decision makers understand as the exact intent of the request for proposals is essential. While it is still far too early to pass judgement on the success of the ASSIST program, we are pleased that our recommendations appear to offer significant advantages over other solutions available, and that they have consequently been implemented by NCI.

**APPENDIX**

The model actually solved has 23 variables and 14 constraints, and can be written as:

Maximize \( \sum_{j=1}^{23} V(j)X_j \)

subject to

\( \sum_{j=1}^{23} p_jX_j \geq P \)

\( \sum_{j=1}^{23} b_jX_j \leq B \)

\( \sum_{j=1}^{23} a_{ij}X_j \geq 3, \; i = 1, \ldots, 4 \)

\( \sum_{j=1}^{23} a_{ij}X_j \geq 2, \; i = 5, \ldots, 12 \)

\( X_j = 0 \) or \( j = 1, \ldots, 23, \)

where

\( X_j = 1 \) if proposal \( j \) is funded, \( X_j = 0 \) otherwise;

\( V(j) \) = the rank function score of proposal \( j; \)

\( p_j \) = the budget preference score of proposal \( j; \)

\( b_j \) = the budget of proposal \( j, \; j = 1, \ldots, 23; \)

\( a_{ij} = 1 \) if proposal \( j \) is from a state in region \( i, \)

= 0 otherwise, \( i = 1, \ldots, 4; \)

\( a_{ij} = 1 \) if proposal \( j \) is from a state in smoking preference quartile \( (i - 4), \)

= 0 otherwise, \( i = 5, \ldots, 8; \)

\( a_{ij} = 1 \) if proposal \( j \) is from a state in decline in smoking rate quartile \( (i - 8), \)

= 0 otherwise, \( i = 9, \ldots, 12; \)

\( P \) = the total preference function requirement;

\( B \) = the available budget.

**ACKNOWLEDGMENT**

LINDO is a product of Lindo Systems, Inc., Chicago, Illinois. We are grateful for the advice given by Professor Linus Schrage (University of Chicago) and Kevin Cunningham (LINDO Systems, Inc.) during the development of the project. We also appreciate the cooperation and encouragement of decision makers at the National Cancer Institute. In particular, we are especially grateful to Drs. Jessie Gruman and Thomas Glynn for considerable advice about the decision criteria. Three referees and an associate editor provided extremely helpful, detailed comments on an earlier draft of this paper.

**REFERENCES**


