A SURVEY BUDGETING SYSTEM

Brian James, consultant, Karol Krotki and Amy Rathbun, Education Statistics Services Institute.
Brian James, 1235 Jeanette Avenue, Union, NJ 07083

Key words: Survey budgeting, Survey management, Computer applications, CATI

1. Introduction

Survey budgeting and costing are central and pervasive activities in survey practice. In particular, cost modeling is an important item of methodology research in large, multi-stage sample surveys like the Current Population Survey (Bryant and Weidman, 1987) and the National Health Interview Survey (Judkins and Waksberg, 1990). These reports give detailed accounts of how cost models can be specified and estimated. In addition, the classic textbooks (e.g. Cochran, 1977) describe interrelationships between cost, variance and sample size under various sample designs. But this body of conceptual and methodological work, while impressive, is still small compared to that for, say, sampling, estimation, questionnaire design, or hypothesis testing. Neither has it spawned a set of standardized computer software applications. Indeed, it appears that each survey research organization has its own proprietary approach to survey budgeting and costing. To the best of our knowledge, there is no generic, off-the-shelf survey budgeting software that is user-friendly, readily available, and general enough to be applicable to a variety of situations.

This paper reports on the early results of a project whose objective was to develop such a system for the National Center for Education Statistics. The system is being developed by the Education Statistics Services Institute and is called the ESSI Survey Budgeting System (SBS).

The spreadsheet-based SBS has been created to produce budget estimates for CATI surveys. With this system, a survey or project manager enters various survey design and cost parameters to determine the projected budget for a survey.

Because the SBS is interactive, it allows a project manager to observe how changes in certain variables, such as target sample size or specific administrative and field costs, will affect overall survey costs and other factors.

The system operates under Microsoft Excel, version 7. The user does not need to be familiar with spreadsheet operations to utilize the system. Even the programming is straightforward and allows for changes in format and content.

In the next section we describe the conceptual framework on which the system is based. Next, we describe the functionality elements which were defined for the system. Section 4 discusses the details of how the system works, form by form. We conclude by outlining the next steps for this project, and suggest a possible path which survey budgeting can follow.

2. Conceptual Framework

The effective sample size achieved in a survey is a random variable whose value is determined by the survey design, the characteristics of the population under study, and the efficiency of the survey organization. Hence, sampling rates, incidence rates for target households, the propensity of target households to cooperate, interviewer success rates, and calling center throughput all help to determine how much sample gets to the tabulation and analysis stage.

The cost of obtaining a given effective sample size, being a function of this sample size, is itself a random variable. A major challenge for survey management is therefore to make reasonable predictions of survey cost, and then to monitor and adjust these estimates as the survey progresses.

Although the ESSI Survey Budgeting System was designed to produce budget projections for telephone surveys conducted by the National Center for Education Statistics, the design parameters are broad enough to accommodate a wide range of telephone-based surveys.

2.1 Purpose

The broad purpose and objectives of the SBS were defined are as follows.

1. The system will be used by survey managers to prepare budgets for upcoming surveys.

2. Managers will be able to update their cost projections on an ongoing basis by entering current sample performance data as these become available.

3. In addition, it was decided that the system should be easy to use, PC-based, and require a minimal amount of training.

3. In addition, it was decided that the system should be easy to use, PC-based, and require a minimal amount of training.

2.2 The Variable Cost Model

We assumed a telephone-based (RDD) sampling methodology with data collection by means of CATI. We also chose to focus on the survey's variable costs, as these tend to consume the largest share of the budget and are the most difficult to predict. For our purposes, variable costs are defined as those costs which vary directly with sample size and/or length of the project.

We need to distinguish between the activities of budgeting and costing. Survey costing provides a basis for survey budgeting. It involves specifying and estimating relationships between various cost items - salaries, traveling, telephones, etc. Survey budgeting involves finding the most cost-effective combination of resources (staff and equipment) to achieve a given level of output (completed questionnaires) with a given level of quality (bias and precision), subject to constraints of time and money. Budgeting assumes a given set of cost relationships.

We also note an absence in the survey budgeting and costing literature of a universal definition for cost as a construct, and for particular subcategories such as fixed, variable, overhead, and even field cost. Such definitions being standard in cost accounting and operations research, for example.

Three types of factors impact a survey's variable costs - **design factors**, **external factors**, and **operational factors**. Figure 1 is a schematic of the survey process showing how sample flows through the system and is impacted by these factors. This is the underlying model for the budgeting system. A description of the process follows.

Design factors include the choice of survey population, sampling frame, desired precision, sample size, data collection method, time schedule, predicted performance rates, etc. These factors determine the initial estimate of how much sample will be selected. Selected sample (or starting sample) is composed of raw telephone numbers, hundred blocks, etc. Some of the subsets of selected sample are: households, businesses, governmental, educational, military and other institutions, out of service , and unassigned.

As selected sample is fed through the CATI system, it is reduced by the effect of external factors such as incidence rates for the target group (e.g. the number of households with a school age child per 1,000 telephone households), and cooperation rates (e.g. the percentage of eligible households in the target group who agree to complete the interview).

Operational factors further reduce the volume of sample-in-progress as it moves through to the stage of completed agrees (i.e. no further call backs, and verified by a supervisor). These include performance factors specific to the data collection organization and to the time of day or the day of week. For example, the rate at which the automatic dialer puts through calls each hour; the rate at which these calls result in a completion (an eligible respondent answers the telephone); and the number of interviews completed by interviewers each hour.

During editing and data cleaning, the number of completed agrees is further reduced resulting in the final count of effective sample used for tabulation and analysis.

The process includes a feedback loop through which actual performance statistics are used to revise earlier budget projections. This allows revised budgets to be generated as needed. It also requires that a complete set of sample performance statistics and unit cost data be maintained by the survey organization.

2.3 Validity and Reliability

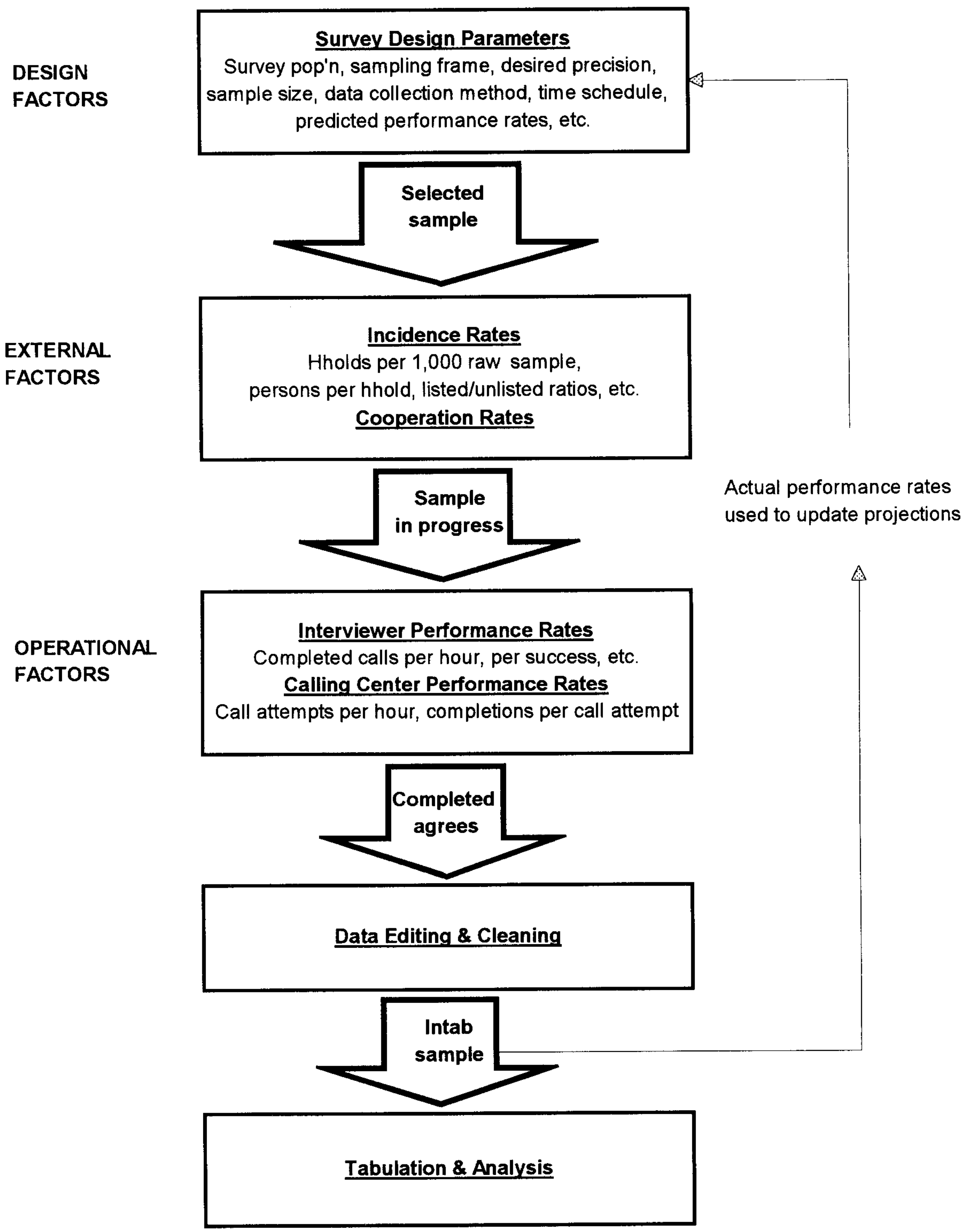
We believe the generalized model outlined above provides a valid description of survey practice as implemented in most CATI-based survey research organizations. However, the validity and reliability of budget projections output by the system depends to a large extent on the quality of sample performance parameters and assumptions entered by the user. User input for any given project should be therefore be based on both historical and current performance data for similar projects.

3. System Functionality

Based on the variable cost model, the following functional requirements were specified for the system.

1. Accept a set of input parameters including:
- a project time schedule broken down by activity, e.g. survey design, staff recruitment, data collection, data processing, report preparation;
- sample design factors, e.g. sample targets, average interview length, etc., broken down by strata or subpopulation;
- projected sample performance factors, e.g. incidence rates, response rates, etc.

Figure 1 - FACTORS AFFECTING VARIABLE SURVEY COSTS



- CATI factors e.g. number of calling centers, workload distribution across centers, paid hours and calling hours per shift, interviewer/supervisor ratios, etc.;
- unit costs for field and office personnel (including fringes), telephone time, equipment, printing, postage, etc.

2. Calculate the amount of starting sample to be selected, the total interviewing hours, number of interviewers and supervisors required at each calling center, number of calling days required, and the variable cost of data collection.

3. Provide budget summaries for any given project, broken out by:
   - time period (week, month, year)
   - budget item and category (staff , telephone, equipment, etc.)
   - calling center.

4. Accept cost factors for editing, tabulation, analysis and publication, and calculate the total cost (fixed and variable) for the entire project.

5. Provide comparisons of actual vs. projected budgets, showing variances.

6. Provide budget summaries across projects, in any of the above formats.

4. Implementation

It was decided to implement the system in two phases, Phase 1 covering the first three functions listed above, and Phase 2 covering the remaining three.

It was also decided to implement the system on a PC platform using a popular spreadsheet program. These packages permit rapid prototyping and modification, have robust user interfaces, provide good graphing and charting capabilities, and can exchange data with other types of software.

The system has been built as a set of linked spreadsheet forms (or templates) requiring the user to input various performance parameters and cost factors. Each form retrieves calculated values stored in the preceding forms, and in turn, later forms access its calculations. Every form is an Excel workbook containing four identical sheets, which for four types of survey activity - two pretests, a main survey, and a follow-up survey.

To date, Phase 1 has been completed and is a working system that produces budget projections. We created two variants of the program. They differ primarily in their use of an explicit time schedule, and also in the calculation of total call duration.

The first variant requires the user to enter the project time schedule under five major activity categories - survey design, recruitment of project staff and vendors, data collection, data processing and report preparation. The user indicates the number of total and non-overlapping days allocated to each activity. These allocations are used in subsequent forms as constraints for various cost items.

This version also works with a detailed breakdown of call dispositions. It requires the user to estimate the percentage distribution of total call attempts across the disposition categories, the average number of call attempts in each category, and the average duration per call in each category. Categories include original completed interviews, refusal conversions, interview refusals, screener refusals, non-households, wrong numbers, busy/no answers, etc.

The second variant (described below) endogenously determines the time spent on each major survey activity, based on other information supplied by the user. For example, the number of data collection days is calculated from the sample size and average interview time. Later, the user enters a factor which relates the days worked by data processing staff to data collection days.

This version of the program allows the user to enter the average interview time, to which is applied a contact rate adjustment factor reflecting the relative difficulty of contacting a household. This factor accounts for time spent cycling through unusable numbers and refusals to get a responding household. Factors can vary at the stratum/subpopulation level.

Following is a brief summary of how each form works. Forms with an 'a' suffix indicate they are different in the two versions of the system. Forms without a suffix are the same in both versions.

Form 1 - Survey Design & Performance Parameters

This form calculates the amount of starting sample (phone numbers) needed to achieve the effective sample target. Starting with the effective target, we work backwards, applying various projected sample performance factors to arrive at the required starting sample. Sample targets are normally stated in terms of the unit of analysis, which is usually persons. Dividing by the average persons per household we convert a persons target to a household target.

Form 2a - Call Duration

This form calculates total calling time needed to complete the sample, and its distribution by calling center. Form 2a is based on a simplified method compared to the earlier version. The user enters an average interview time and a contact rate adjustment factor and the product of these is multiplied by the number of target interviews to arrive at total interview time.

Form 3a - Unit Cost & Total Cost

This form calculates the required numbers of interviewers, supervisors, paid interviewer hours, paid supervisor hours, and calling days. Form 3a also has an earlier version. The user enters the hours per shift, shifts per day, and interviewer/supervisor ratios, which, when combined with call center workload distributions from Form 2a, result in required numbers of interviewers and supervisors. We distinguish between paid hours per shift and calling hours per shift, the former including down time for lunch breaks, etc. These factors help to determine how many interviewers are required per shift as well as the required number of calling days.

Form 4 - Unit Cost & Total Cost

This form calculates an itemized total cost for the project and sums up the resource utilization in all four phases of the project. The basic relationship is "Price * Quantity = Total Cost". For each item, the user enters a unit of measurement (hours, days, etc.), the cost per unit, and the quantity of units to be utilized in each phase of the project (certain quantities, such as number of interviewers, are automatically retrieved from earlier calculation). These are multiplied out to get total cost per phase, then summed to get the Project Total. The items are broadly grouped into Field Costs and Administrative Costs. Within these groups are categories - field staff, training, etc.- and within these, individual items

Form 5 - Allocation of Costs by Month

This form estimates the total project cost on a month by month basis. The two sheets, "Allocation - %" and "Allocation - $", show the percentage allocation of costs by month, and the equivalent dollar amounts, respectively. In the "Allocation - %" sheet, the user enters the starting month and year for the project, and a series of monthly percentage values, for each cost item. The "Allocation - $" sheet shows the result of applying these percentages to the project totals in Form 4. It is entirely calculated and requires no user input.

Form 6 - Projected Budget - Summary by Item

This form summarizes the budget projections for each item, in user-defined time increments. The user specifies the staring and ending month and year for a summary period, e.g. Jan 1997 to Mar 1997. The program then sums the corresponding months from the Allocation - $ sheet (Form 5) and enters the total in a column with the heading specified. Summaries for up to eight user-defined time periods can be generated together.

Form 7 - Projected Budget - Summary by Category

This form summarizes the budget projections for each cost category, in user-defined time increments. It is similar to Form 6 except that costs are summarized for categories (Field Staff, Training, etc.) instead of individual items.

5. Conclusions

The ESSI Survey Budgeting System, provides an effective and uniform means of preparing and updating survey budgets.

Future improvement and expansion will include: adding the last three functions listed in section three, and in particular, analysis of the variance between actual and projected budgets, enhancing the user interface, and incorporating other survey methods such as mail and CAPI.

As far as the discipline of survey budgeting and costing is concerned, it may well be that rapid methodological development will occur and standardized tools will emerge only when survey researchers adopt an interdisciplinary approach. Cost accounting and operations research are two disciplines which can offer powerful tools and techniques for improving current survey budgeting practices.

References

Bryant, K.J., Weidman, L. (1987). Developing cost models for CATI surveys. *ASA Proceedings of Survey Research Methods Section*, 767-771.

Cochran, W.G. (1977). *Sampling Techniques (3rd Ed.).* New York: John Wiley & Sons.

Judkins, D., Waksberg, J., Northrup, D. (1990). Cost functions for NHIS and implications for design. *ASA Proceedings of Survey Research Methods Section*, 34-43.

Kalsbeek, W.D., Botman, S.L., Massey, J.T., Liu, P. (1994). Cost-efficiency and the number of allowable call attempts in the National Health Interview Survey. *Journal of Official Statistics*, 10, 133-152.