A SYSTEMS APPROACH TO DETERMINING THE COST AND COMPARING ALTERNATIVE SURVEY/CENSUS DESIGNS

WILLIAM F. MCCARTHY
U.S. DEPARTMENT OF TRANSPORTATION,OFFICE OF THE INSPECTOR GENERAL,WASHINGTON DC
UNIVERSITY OF MARYLAND,UNIVERSITY COLLEGE GRADUATE SCHOOL,COLLEGE PARK MD

INTRODUCTION

This paper presents a systems approach to determine the cost of a survey/census design. This approach allows the analyst to determine the cost in a reliable and accurate manner, to compare alternative designs, and to determine the practicability of each design. This approach uses a system cost element list (Fisher, G.H.,1970A; Fisher, G.H.,1970B; and Massey, H.G, D. Novick, and R.E. Peterson, 1972) and a technique which allows for a meaningful comparison of alternatives (known as normalization - Fisher, G.H., 1970A).

The first section of the paper will describe how the analyst defines a survey/census design as a system. The second section describes how to determine the cost of a survey/census design. The third section describes how to compare alternative designs. The fourth section presents other applications of a system cost element list. Finally, the fifth section presents an example of how a system cost element list is used.

1. DEFINING A SURVEY/CENSUS DESIGN AS A SYSTEM

Definition 1. In general, a survey/census design involves the following aspects:

A. The survey/census objectives.
- i. the definition of the variables
- ii. the specification of the population
- iii. the method(s) of data collection
- iv. the method(s) of data processing
- v. the method(s) of data analysis
- vi. the specification of the desired precision/accuracy of the estimates/counts

B. The sample/enumeration design.
- i. the selection/coverage process
- ii. the estimation/count process

(Adaptation of Kish,L., 1965).

The operational and support activities associated with a particular survey design are listed in TABLE 1 and those associated with a particular census design are listed in TABLE 2.

Definition 2. In general, "a system is a set of units or elements which are actively interrelated and which operate in some sense as a bounded unit" (Baker,F., 1973).

In terms of this paper, this bounded unit is nothing more than the complete collection of operational and support activities which are associated with a particular survey/census design. Therefore, the collection of activities listed in TABLE 1 and TABLE 2 can be considered a systems definition of a survey design and a census design, respectfully.

2. HOW TO DETERMINE THE COST OF A SURVEY/CENSUS DESIGN

A system cost element list can be considered a general framework within which the analyst delimits the aggregate of operational and support (O&S) activities in terms of their corresponding costs and overheads. This framework provides that all costs associated with a survey/census design are identified and included (referred to as inclusiveness).When considering alternative designs, this framework not only ensures inclusiveness, but also consistency.By consistency, one means that the analyst uses the same types and categories of cost for each alternative and, whenever possible, the same or similar data sources and derivation methods to develop the O&S activity requirements and their associated costs and overheads. TABLE 3 illustrates the framework of a system cost element list. TABLE 4 illustrates how this framework can be applied to a survey (in terms of field O&S activities).

3. HOW TO COMPARE ALTERNATIVE DESIGNS

Not only is inclusiveness and consistency important when considering alternative designs, but also the analyst should be confident that the framework also provides for comparability.As noted by Fisher(1970A), in order to develop a comparative framework which produces meaningful comparisons, the analyst needs to "normalize" the alternatives with respect to either effectiveness (utility, benefit) or cost (total budget available). This process of normalizing the alternatives ensures that the analyst ultimately compares "apples" with "apples". As noted by Fisher, "In any event, the objective is to permit comparisons to be made among alternatives, and for this purpose something has to be made fixed" (1970B, p11.).

In making such comparisons, with the ultimate goal of selecting the "best" design(s), the analyst can make use of two "normalization" approaches:

1. Fixed Effectiveness Approach

For a specified level of effectiveness to be attained in the accomplishment of some given objective, the analyst attempts to determine that alternative (or feasible combination of alternatives) which is likely to achieve the specified level of effectiveness at the lowest economic cost.

2. Fixed Budget Approach

For a specified cost level to be used in the attainment of some given objective, the analyst attempts to determine that alternative (or feasible combination of alternatives) which is likely to produce the highest effectiveness.

The analyst can define "effectiveness" in a number of ways.However, it is difficult to make judgements about the relative utility or worth of the various alternative designs. These judgements can be qualitative in nature. Therefore, the fixed effectiveness approach may not ensure an exact comparability between the alternative designs (comparability can only be demonstrated in a general "fuzzy" sense). The definition of a budget is more straight forward. Due to the fact that the specified cost level to be used is quantitative in nature, one can expect exact comparability between the alternative designs when using the fixed budget approach. The author has tried to quantify "effectiveness" by simply considering relative utility or worth of a design in terms of statistical precision or accuracy.If the fixed effectiveness approach is used, each alternative design would have the same specified level of statistical precision or accuracy. Thus, the alternative(s) with the lowest economic cost (total budget) would be considered for implementation. The author has used this approach when comparing alternative survey/ census designs (e.g., McCarthy, W.F.,1988; McCarthy, W.F., 1989; and Herriot,R., D.V. Bateman, and W.F. McCarthy, 1989).

The fixed approach for comparing alternative designs can be misused. For example, if the analyst is interested in determining if a new census design (say, one that utilizes a new data collection approach) is less expensive than the current census design being used, he/she might be tempted to normalize the designs with respect to "the number of household to be enumerated". If the new census design generated a lower cost (requires a lower total budget), the analyst might believe that this new design should replace the design currently being used.This could be a mistake! The statistical accuracy of the new census design (based on coverage) may not be identical to that generated by the current design. If it is important that the census results attain a certain level of statistical accuracy (which the current design produces) and the new design in reality does not produce an identical level of statistical accuracy, then the selection of the new design could be disastrous. Therefore, it is important that the analyst does not incorrectly normalize the alternative designs. If the normalization is incorrect, then the comparative analysis analysis will produce meaningless results (i.e., the analyst in reality would be comparing "apples" with "oranges".

The fixed approach could also produce meaningless results if the data sources being used to determine the cost of the various alternatives are not similar (i.e., we have a lack of consistency with respect to data sources). An example of this would arise if the data source for one alternative was based on empirical values and the data source for the other alternative under consideration was based on non-empirical (a priori) values.

There are two techniques available to the analyst which methodically deal with such a case of inconsistency and help to still ensure a meaningful comparison. One technique is referred to as a <u>sensitivity analysis</u> and the other is referred to as an <u>a fortiori analysis</u>.

The analyst can use a sensitivity analysis to compare the alternative design based on non-empirical (a priori) information to the other design which is based on firmer empirical information. (Note: To simplify the narrative, the design based on non-empirical [a priori] information will be referred to as alternative 1; the design based on the firmer empirical information will be referred to as alternative 2; and the analyst is using the fixed effectiveness approach.)

Essentially the analyst would see how alternative 1 compares with alternative 2 when the non-empirical (a priori) values of alternative 1 are incrementally increased and/or decreased. This incremental increasing/ decreasing can be done in such a way as to generate the following value categories: "<u>optimistic</u>", "<u>average</u>", and "<u>pessimistic</u>".

The next step would be for the analyst to run the different value categories separately through alternative 1 (with respect to the system cost element list), thus generating a total cost for each value category.The analyst would then compare each of the total cost values to the one generated by alternative 2. If alternative 1 generates the lower total cost under the pessimistic category when compared to alternative 2, the analyst has a strong case for selecting alternative 1 over alternative 2. If however, alternative 1 only generates a lower cost under the average and/or optimistic categories, the analyst is taking a risk if he/she selects alternative 1 over alternative 2. In all likelihood, the analyst would be better off selecting alternative 2 under this situation. It goes without saying that if alternative 1 does not have the lower total cost under any value category, that alternative 2 is selected. The author has used sensitivity analysis in this manner when dealing with a lack of consistency with respect to data sources (e.g., McCarthy, W.F., H.Montagliani, and L. McGinn, 1988; McCarthy, W.F., 1989; and Herriot,R., D.V. Bateman, and W.F. McCarthy, 1989).

A fortiori analysis works in a similar manner. When comparing alternative 1 to alternative 2, the analyst deliberately chooses to make the non-empirical (a priori) information used in alternative 1 more optimistic than the firmer empirical data used in alternative 2.If alternative 2 still generates the lower total cost, the analyst has a very strong case for selecting alternative 2 over alternative 1.

The fixed approach with respect to a system cost element list also allows the analyst to see whether or not the various O&S activities of each alternative design exceed their particular constraints (i.e., the analyst is able to, in a meaningful manner, perform a "<u>sanity test</u>" in order to see if the various alternative designs are practicable in terms of staff requirements, the number of interviews to be conducted, the amount of time it takes to complete an interview, the amount of travel required, etc.). This additional advantage can allow the analyst to not only select the "best" alternative design(s) but also to make sure that the "best" design(s) are practicable as well. Selecting a design that has the lower cost (or higher effectiveness) without considering whether or not it the design(s) is practicable, can be very dangerous!

If several designs are considered the "best", the analyst could use this "sanity test" to rank the "best" designs, and ultimately select one over the others. In any event, it is wise that the analyst performs this sanity test before the final selection of a design is made.

4. OTHER APPLICATIONS OF THE SYSTEM COST ELEMENT LIST

The analyst can use the system cost element list with respect to an existing (on-going) survey/census design in order to make sure that all costs are identified and included (inclusiveness). Once the analyst is satisfied that all costs are identified and included, he/she can then determine, in a reliable and accurate manner, what the variable and fixed (constant) costs are associated with that particular design. These variable and fixed costs can then be incorporated into a cost function. A cost function essentially summarizes the great variety of expenditures (variable and fixed) of a survey/census design (Refer to Hansen, M.H., W.H. Hurwitz, and W.G. Madow, 1953 or Kish, L., 1965 for more information about cost functions and how they are used).

Also, the analyst can develop a system cost element list for each existing survey/census design. When a new survey/census design is considered, the analyst could match the new survey/census design to a similar existing design. Once this match has been made, the analyst could then estimate by analogy what the O&S activities would be for the new design as well as its variable and fixed costs.

5. AN EXAMPLE OF HOW THE SYSTEMS APPROACH IS USED

Since 1981, the U.S. Bureau of the Census has been conducting studies to evaluate Computer Assisted Telephone Interviewing (CATI) in order to see if it is a feasible methodology for conducting surveys. Recently, two studies conducted by the author, have looked at the use of CATI. One, Study A, looked at the use of CATI as a partial means of data collection for the National Crime Survey (NCS), a current demographic survey (McCarthy, W.F. et al, 1988). The other, Study B, looked at the use of CATI as the major means of collecting census sample data (Herriot, R. et al,1989).

This section will briefly outline how a systems approach was used to compare the costs of using CATI in conjunction with the current data collection methodology used to collect survey/census data (primarily personal interviews or mailback questionnaires) against the costs of only using the current data collection methodology.

STUDY A

A system cost element list was developed to analyze the production requirements and associated costs for two alternative designs: Alternative A- conducting a demographic survey with CATI in conjunction with the current data collection methodology and Alternative B - conducting a demographic survey with only the current data collection methodology.

STUDY B

A system cost element list was developed to analyze the production requirements and associated costs for two alternative designs: Alternative A - conducting a census sample survey primarily with CATI (households without telephones would be contacted via personal interview and/or mail) and Alternative B - conducting a census sample survey with only the current data collection methodology.

Basically, the production requirements entail the workload distribution (the average number of interviews (cases) assigned to CATI or the current data collection methodology), the staff distribution (the number of interviewers, supervisors, etc., that are required to meet the demands of some specified workload distribution), time constraints (the amount of time available to conduct a particular survey), the salary and benefit requirements of the staff, and assumptions about the unit costs for interviewed and non-interviewed cases (for both CATI and the current data collection methodology), and other associated fixed and variable costs for such things as sample design and selection, development of training materials (for the staff), telephone costs for both CATI and the current data collection methodology, questionnaire development and printing, data keying, programming, data analysis, publication of results, etc.

The resulting costs that are generated were in the identical format of the cost reports put out by the Census Bureau's Budget Office. These cost reports, in addition to other reports, are used by the Bureau's decision-makers to keep track of and to analyze the costs associated with conducting surveys.

In general, both studies used the framework illustrated in TABLE 4. Once the system cost element list for a study was developed on paper, it was translated into a spreadsheet format via LOTUS 1-2-3. The spreadsheet format made the manipulation of the system cost element list easier (e.g., when performing sensitivity analyses). Both studies were normalized with respect to fixed effectiveness (statistical precision).

SUMMARY

A systems approach can be helpful to the analyst if he/she is interested in determining the cost of a survey/census design in a reliable and accurate manner. In addition, this systems approach allows the analyst to compare alternative designs in a meaningful manner as well as to determine the practicability of each design.

REFERENCES

Baker,F. (1973), Organizational Systems: General Systems Approaches to Complex Organizations, Richard D. Irwin, Inc., Homewood, IL.

Fisher, G.H. (1970A), Some Comments on Conceptual Frameworks for Comparing Alternatives, Paper P-4506, Rand Corporation, Santa Monica, CA.

Fisher, G.H. (1970B), Cost Considerations in Systems Analysis, American Elsevier Publishing Co., Inc., New York, NY.

Hansen, M.H., W.H. Hurwitz, and W.G. Madow (1953), Sample Survey Methods and Theory, Vol.1, John Wiley & Sons, New York, NY.

Herriot,R., D.V. Bateman, and W.F. McCarthy (1989), A Decade Census Program-A New Approach for Meeting the United States of America's Needs for Sub-National Data, Proceedings of the Social Statistics Section, American Statistical Association.

Kish, L. (1965), Survey Sampling, John Wiley and Sons, New York, NY.

Massey, H.G., D. Novick, and R.E. Peterson (1972), Cost Measurement: Tools and Methodology for Cost Effectiveness Analysis, Paper P-4762, Rand Corporation, Santa Monica, CA.

McCarthy, W.F., H. Montagliani, and L. McGinn (1988), NCS/CATI Cost Analysis - Final Report, Unpublished CATI Research Report, No. NCS-1, U.S. Bureau of the Census.

McCarthy, W.F. (1988), OR/MS Applications in Computer Assisted Telephone Interviewing (CATI) Survey Research; Part1: Cost Modeling, paper presented to The Institute of Management Sciences and Operations Research Society of America, Joint National Meeting, April, Washington, DC.

McCarthy, W.F. (1989), Evaluation of Computer Assisted Telephone Interviewing as a Survey Methodology by Means of Cost Models and Mathematical Programming, in Cost Analysis Applications of Economics and Operations Research, edited by T.R. Gulledge and L.A. Litteral, Springer-Verlag, New York, NY.

TABLE 1
MAJOR O&S ACTIVITIES ASSOCIATED WITH A SURVEY

(in no particular order)

* INITIAL PLANNING
* SAMPLE DESIGN
* SAMPLE FRAME CONSTRUCTION/MAINTENANCE
* SAMPLE SELECTION
* QUESTIONNAIRE DEVELOPMENT AND PRODUCTION
* PROCEDURES/SPECIFICATIONS DEVELOPMENT
* RECRUITMENT OF STAFF
* TRAINING MATERIAL DEVELOPMENT AND PRODUCTION
* TRAINING OF STAFF
* QUALITY ASSURANCE/CONTROL
* PRE-TESTING
* PUBLIC AWARENESS
* SUPERVISION
* DATA COLLECTION
* TRAVEL
* DATA EDITING
* DATA PROCESSING
* DATA ANALYSIS
* POSTAGE
* PUBLICATION OF RESULTS
* DISSEMINATION OF PUBLICATION

TABLE 2
MAJOR O&S ACTIVITIES ASSOCIATED WITH A CENSUS

(in no particular order)

* INITIAL PLANNING
* CENSUS DESIGN
* MAPPING/LISTING
* QUESTIONNAIRE DEVELOPMENT AND PRODUCTION
* PROCEDURES/SPECIFICATIONS DEVELOPMENT
* RECRUITMENT OF STAFF
* TRAINING MATERIAL DEVELOPMENT AND PRODUCTION
* TRAINING OF STAFF
* QUALITY ASSURANCE/CONTROL
* PRE-TESTING
* PUBLIC AWARENESS
* SUPERVISION
* ENUMERATION (DATA COLLECTION)
* TRAVEL
* POSTAGE
* DATA EDITING
* DATA PROCESSING
* DATA ANALYSIS
* PUBLICATION OF RESULTS
* DISSEMINATION OF RESULTS
* POST ENUMERATION EVALUATION

TABLE 3
THE GENERAL FRAMEWORK OF A SYSTEM COST ELEMENT LIST

ACTIVITIES	ALTERNATIVES			
	A	B	C	...
ACTIVITY 1				
COST ELEMENT 1	X	X	X	...
COST ELEMENT 2	X	X	X	...
.				
.				
.				
COST ELEMENT N	X	X	X	...
ACTIVITY 2				
COST ELEMENT 1	X	X	X	...
COST ELEMENT 2	X	X	X	...

TABLE 3 CONTINUED

ACTIVITY 2	A	B	C	...
.				
.				
COST ELEMENT N	X	X	X	...
ACTIVITY N				
COST ELEMENT 1	X	X	X	...
COST ELEMENT 2	X	X	X	...
.				
.				
.				
COST ELEMENT N	X	X	X	...
OVERHEADS				
OVERHEAD 1	x	x	x	...
OVERHEAD 2	X	X	X	...
.				
.				
.				
OVERHEAD N	X	X	X	...
TOTAL COST	Y	Y	Y	...

WHERE X = SOME DOLLAR VALUE BASED ON EMPIRICAL AND/OR A PRIORI INFORMATION (i.e., A DOLLAR VALUE BASED ON A COST ESTIMATING RELATIONSHIP, A COST FUNCTION, AN ANALOGY ESTIMATE, ETC.: Y= THE SUMMATION OF ALL COST ELEMENTS AND OVERHEADS ASSOCIATED WITH A PARTICULAR ALTERNATIVE.

TABLE 4
THE FRAMEWORK APPLIED TO A SURVEY

PRODUCTION REQUIREMENTS	ALTERNATIVES	
	A	B
AVG. NO. CATI TEL. CASES	C	C
AVG. CASELOAD PER INTERVIEWER	C	C
AVG NO. NONINTERVIEW CASES FOR CATI	C	C
AVG. NO. PERSONAL INTERVIEW CASES	C	C
AVG. CASELOAD PER INTERVIEWER	C	C
AVG. NO. NONINTERVIEW CASES FOR PERSONAL INTERVIEWS	C	C
AVG. INTERVIEW TIME PER CATI CASE	C	C
AVG. INTERVIEW TIME PER CATI NONINTERVIEW CASE	C	C
AVG. INTERVIEW TIME PER PERSONAL INTERVIEW CASE	C	C
AVG. INTERVIEW TIME PER PERSONAL NONINTERVIEW CASE	C	C
AVG. TRAVEL PER PERSONAL INTERVIEW CASE	C	C
AVG. TRAVEL PER PERSONAL NONINTERVIEW CASE	C	C
AVG. COST PER CATI INTERVIEW CASE	C	C
AVG. COST PER CATI NONINTER-VIEW CASE	C	C
AVG. COST PER PERSONAL INTERVIEW CASE	C	C
AVG. COST PER PERSONAL NONINTERVIEW CASE	C	C
TRAINING COST PER STAFF MEMBER	C	C

TABLE 4 CONTINUED

	A	B
MILEAGE RATE PER MILE	C	C
STAFF SALARIES		
COORDINATORS	C	C
PROGRAM SUPERVISOR	C	C
CLERKS	C	C
INTERVIEWERS	C	C
PERSONNEL REQUIREMENTS		
NO. COORDINATORS	C	C
NO. PROGRAM SUPERVISORS	C	C
NO. OF CLERKS	C	C
NO. OF INTERVIEWERS	C	C
FIELD COSTS		
INTERVIEWING	X	X
REINTERVIEWING	X	X
TRAINING	X	X
OFFICE WORK	X	X
SUBTOTAL A: TOTAL DIRECT COST	Z	Z
APPLICATIONS & BENEFITS		
PERSONAL BENEFITS	X	X
ETC.	X	X
SUBTOTAL B: TOTAL PERSONAL BENEFITS	Z	Z
OVERHEADS	X	X
SUBTOTAL C: TOTAL OVERHEADS	Z	Z
TOTAL FIELD COST	Y	Y

WHERE C= SOME FIXED VALUE; X = SOME VALUE GENERATED BY A FORMULA; Z =A SUMMATION OF COST ELEMENTS WHICH MAKE UP A SUBTOTAL; AND Y= THE SUMMATION OF ALL SUBTOTALS.

FOR A MORE COMPREHENSIVE ILLUSTRATION OF AN APPLICATION OF THE FRAMEWORK TO A SURVEY, PLEASE REFER TO McCarthy, W.F., et al.1988.