

COST MODELS FOR COMPUTER-ASSISTED INTERVIEWING IN THE CPS

William L. Nicholls II, Michael E. Haas, and Richard F. Blass, U.S. Census
W. Nicholls, Field Div., Census Bureau, Dept. of Commerce, Washington D.C. 20233

KEYWORDS: Survey costs, cost model, CATI, CAPI

The U.S. Census Bureau and the Bureau of Labor Statistics (BLS) are planning to replace paper-and-pencil with computer-assisted interviewing methods in the Current Population Survey (CPS). This paper describes cost modeling activities undertaken to estimate the data collection costs of alternative scenarios for the utilization of computer-assisted telephone and personal interviewing (CATI and CAPI) in the CPS.

1. CPS Design and CATI-CAPI Scenarios

1.1 Design of the Current CPS

The CPS is conducted primarily to estimate employment, unemployment, and other characteristics of the civilian noninstitutionalized population of the United States. The national sample is composed of independent samples for the 50 states and the District of Columbia and consists of approximately 67,300 addresses per month, currently reduced from its traditional 71,300 addresses per month for budgetary reasons.

Each month's sample is composed of eight panels which rotate on a schedule of 4 months in, 8 months out, and 4 months in before sample retirement. Information is collected by personal visit interviews in the first and fifth month in sample (MIS) and by telephone interviews in MIS 2-4 and 6-8 when possible and acceptable to the respondent. About two-thirds (65.5 percent) of all CPS interviews are completed by telephone. These calls are currently made by the field interviewers from their own homes.

The CPS labor force questions are asked each interview month about each applicable household member and refer to activities during the week containing the 12th of the month. Interviewing takes place during the week of the 19th and must be completed by the following Monday. Supplements are added to the base CPS interview in many months to cover such topics as school enrollment, birth expectations, and voting and registration. Most add no more than 4 minutes to the average 10-minute base interview; but the annual March supplement on work experience, income, and migration roughly triples CPS interview length.

Various scenarios have been proposed to introduce CATI and CAPI into CPS data collection. We will focus on two.

1.2 The 5,500 Case CATI Scenario

The first is the transfer of approximately 5,500 CPS telephone interviews per month to CATI while continuing paper-and-pencil data collection for all other CPS personal and telephone interviews. This is a relatively modest scenario utilizing existing resources. The CPS questionnaire, sample design, and treatment of supplements remains unchanged; and CATI interviews are conducted from the Census Bureau's Hagerstown Telephone Center (HTC) employed for CPS/CATI feasibility and evaluative tests for

more than three years. The 5,500 case limit is the projected capacity of the HTC after expansion from 32 to 60 interviewing stations but allowing for the long March CPS supplement each year.

CATI interviews will be used only in MIS 2-4 and 6-8 of the current design. Personal visit interviews will continue in MIS 1 and 5 to encourage respondent participation and to identify vacant and other ineligible housing units. Personal interviews also will continue in all sample months for households without telephones, those not receptive to telephone interviews, and other special cases. Households unreachable from the HTC or refusing a CATI interview will be recycled to the field for followup and completion. Allowing both for personal interviews in MIS 1 and 5 and cases remaining in the field in other sample months, roughly half the CPS addresses can be assigned to CATI. About 11.0 percent of the CATI cases are recycled to the field.

For the 5,500 Case scenario, cases will be transferred to CATI only from CATI-eligible primary sampling units (PSUs). These are larger, urban PSUs with multiple CPS interviewers. Smaller PSUs are considered CATI-ineligible because loss of their telephone cases to CATI could reduce the interviewers' workloads (and therefore compensation) below levels necessary to retain their services.

Table 1 presents results of the cost model described below that compares estimated data collection costs per case of the 5,500 Case scenario with current CPS costs. Outside the CATI PSUs, costs remain the same. In CATI PSUs, the cost of cases remaining in the field increases slightly because field interviewers have somewhat larger proportions of personal visit cases to complete which are dispersed over a somewhat wider geographic area. Cost-efficient CATI cases conducted from the HTC, however, are estimated to keep total cost per case at or below current levels. Based partly on these estimates, phase-in of this scenario is planned to begin in January 1989 and to attain 5,500 cases with attrition-based procedures in two to three years.

Table 1 -- Estimated Data Collection Costs Per Case for The 5,500-Case Scenario Compared With the Current CPS in FY 1987 Dollars

	Present CPS	5,500 Case Scenario
For all cases	\$20.70	\$20.55
CATI-eligible PSUs	20.60	20.20
At HTC	NA	8.15
In Field	20.60	22.85
CATI-ineligible PSUs	20.80	20.80

1.3 Two-Phase CPS

The Two-Phase CPS is a far more ambitious scenario. Under this plan, the CPS sample will be increased to 116,000 addresses per month and interviews will be conducted over a two-week

(rather than a one-week) period for more efficient use of CATI facilities and CAPI hardware. The first week's sample will provide national labor force estimates while the second augments state samples, where needed, to provide comparable estimates for every state.

The Two-Phase CPS will rely exclusively on CATI and CAPI data collection. The size of field interviewing assignments will be changed to permit transfer of virtually all CPS telephone interviews to CATI. The entire CPS sample becomes CATI-eligible, and two additional CATI facilities will be opened to accommodate the workload. Variations in monthly workloads will be reduced by spreading the long March supplement over three months. Portable microcomputers (CAPI) will be used for all personal interviews and for occasional telephone interviews conducted by the field interviewers. Cases will be transmitted among headquarters, CATI, and CAPI via telecommunications.

The CPS questionnaire also will be revised in part to take advantage of such computer-assisted interviewing capabilities as online editing and dependent interviewing. Implementation of the Two-Phase CPS will require: (1) major changes in the design of the CPS sample; (2) procurement of CATI and CAPI equipment; and (3) major research programs to develop the new questionnaire, to evaluate the effects of two-week (rather than one-week) recall on labor force questions, and to compare estimates obtained with the new and old methods.

Table 2 presents further results from the cost model which compare estimates of the data collection costs of the Two-Phase CPS with those of current methods. In the Two-Phase CPS, the costs of field cases are projected to increase dramatically from \$20.70 to \$34.70 per case for two main reasons: (1) these cases now consist almost entirely of personal visit interviews rather than a mixture of personal and telephone visits; and (2) more geographically dispersed field assignments require more travel time and costs. Nevertheless, the lower unit cost of CATI cases should bring the average survey data collection cost per case below current levels.

**Table 2 -- Estimated Data Collection Costs Per Case For the Two-Phase CPS Compared with the Present CPS in FY 1987 Dollars:
Primary Assumptions**

	Present CPS	Two-Phase Design
For all cases	\$20.70	\$20.55
Field Cases	20.70	34.70
CATI cases	NA	7.30

The astute reader will notice that the unit data collection costs of the 5,500 Case CATI scenario and the Two-Phase CPS are the same: \$20.55 per case. This is a coincidence. The 5,500 Case scenario keeps unit costs low by relying primarily on existing resources, such as the HTC. The Two-Phase CPS keeps costs low by a sample and field work design intended to minimize costs. Intermediate scenarios which require new facilities without changes in survey design tend to drive costs up.

Since the Two-Phase CPS will require departmental, OMB, and Congressional review (and funding), final plans (if approved) may differ in many ways from those described above. If approved as currently proposed, the Two-Phase CPS could be in place by 1996 [6].

Cost estimates for both scenarios are based on an extensive set of explicit assumptions which are not repeated here but have been reported elsewhere [5]. The same report provides estimates of continuing administrative and processing costs and of one-time start-up costs for the Two-Phase CPS.

2. Cost Modeling Activities

Cost modeling activities described in this paper began in 1985 as part of a larger Census Bureau effort to develop, evaluate, and implement computer-assisted interviewing in its current surveys [4]. Bryant and Weidman [1] developed the first CPS/CATI cost model. McCarthy [2,3] contributed to its evolution and has prepared related models for the National Crime Survey and for national data collection options after the 1990 Census.^{2/}

All cost models in this family use LOTUS 1-2-3 spreadsheet functions to itemize estimated or known survey unit costs, aggregated to provide comparative cost estimates for alternative data collection scenarios. The current CPS cost model employs 183 input variables and 185 LOTUS equations to generate its estimates. The model has three main functional components.

2.1 Survey Workload Allocation

The first is an allocation of the survey workload across the types of data collection methods employed. The total sample is partitioned into mutually exclusive and exhaustive subsamples, and each subsample is then repartitioned until final subsamples with relatively homogeneous data collection methods (and unit costs) are reached. A graphic summary of the survey workload allocation is provided by an inverted "tree" diagram, illustrated in Figure 1.

In the LOTUS equations, the workload allocation is conveniently represented by a value for the total sample size and by proportions, p, summing to 1.000, which express the division of the sample at each level. The case numbers, n, can then be quickly revised for different scenarios. The values in Figure 1 are those of the 5,500 Case scenario. To represent a scenario in which all PSUs are CATI-eligible, we set the p in box B equal to 1.000 and the p in box C to 0. The case numbers in all other boxes can be quickly calculated. Similarly, to represent the baseline scenario for current data collection methods without CATI, we can set the p in box F to 0 and that in box G to 1.000.

2.2 Basic Interviewing Costs

The second main functional component of the model is a set of simple equations calculating basic interviewing costs for each subset of the sample with approximately homogeneous data collection methods and unit costs. For personal interviews (and for personally confirmed

noninterviews), the basic costs are interviewers' salaries and travel costs. Let:

C_j = Total basic unit cost of a personal interview in sample subset j.

C_{tj} = Travel unit cost of a personal interview in sample subset j.

C_{sj} = Salary unit cost of a personal interview in sample subset j.

A_{ij} = Adjustment factor i in sample subset j.

Then:

$$C_j = C_{tj} + C_{sj}$$

$$C_{tj} = \left(\frac{\text{miles}}{\text{case}} \right) \left(\frac{\text{cost}}{\text{mile}} \right) (A_{1j}) + \left(\frac{\text{other travel costs}}{\text{case}} \right) (A_{2j})$$

$$C_{sj} = \left(\frac{\text{minutes}}{\text{case}} \right) \left(\frac{1}{60} \right) \left(\frac{\text{staff costs}}{\text{hour}} \right) (A_{3j})$$

In brief, unit travel costs are estimated by multiplying estimated miles per case by costs per mile and adding an estimate of other travel costs per case, such as tolls and parking. Unit staff costs are estimated by interviewer minutes per case converted to hours and multiplied by average hourly staff costs. Because Census Bureau interviewers may be any of three types of employees (intermittent, part-time mixed tour, or supervisory field representatives) with different salary levels and benefits, hourly staff costs must be a weighted sum as follows:

$$\frac{\text{staff costs}}{\text{hour}} = \sum_{k=1}^3 P_{kj} s_k b_k$$

where P_{kj} is the proportion of work type k performs in subsample j, s_k is the hourly salary for type k, and b_k is the benefits factor for type k. Similar equations can be written for telephone interviews by replacing travel with telephone costs per case. For CATI interviews conducted from central sites, evening and weekend salary differentials also may apply.

The model is activity based. The required number of interviewing hours is calculated from the case allocation and estimated minutes per case. The number of interviewers is not an input variable but an output variable and an interim product used in estimating training and related costs. Since Census Bureau interviewers and supervisors frequently work on more than one survey in the same month, the same procedures are followed for supervisory activities.

Many of the equations include adjustment factors, A_{ij} , to simplify sensitivity analyses using the LOTUS DataTable option, as explained more fully in Section 3.3 below.

2.3 Survey Field Support System

The third main functional component of the model represents the field support system. It estimates cost of other activities necessary to prepare for and support the basic interviewing function. For field interviews, the field support system includes interviewer training, reinterviewing, field observation, and other elements shown in Figure 2. For CATI interviews, the support system includes interviewer training, reinterviewing, supervisory monitoring and quality circle meetings.

This part of the model has the largest number of variables and equations; but they involve relatively familiar elements. As illustrated in Figure 3, the module for interviewer training begins with estimates of the number of field interviewers to be trained, taking current turnover into account. Then the number of training hours per interviewer is estimated together with the number of supervisory hours required. Salary and benefit costs for trainees and trainers are estimated by type of employee, and travel and other costs are included where necessary.

3. Recent Cost Modeling Activities

Cost modeling activities in the past year have focussed on the preparation of cost estimates for the 5,500 Case and Two-Phase CPS scenarios, previously presented in Section 1, and on three methodological developments: (1) validation and maturation of the model; (2) methods of model presentation; and (3) sensitivity analysis.

3.1 Validating and Maturing the Model

By "validating" a cost model we mean demonstrating that it yields estimates for current scenarios which approximately equal actual costs known from the accounting system or similar sources. The degree of concurrence required between model estimates and "known" values depends on the way the model is used. If the model is used only to demonstrate that a new scenario costs no more than current methods, it may be sufficient to show that its total costs, when based on known or "worst case" estimates, are below total current costs. Similarly, if the purpose of the model is to determine which of two scenarios is relatively less expensive, it may be possible to build sufficient parallelism into the estimates of the two scenarios' components to reach a conclusion without full validation. However, when the purpose of model building is to compare the costs of alternative scenarios as they are proposed, and to prepare preliminary budget estimates for each, then model validation becomes critical.

We went beyond model validation to "benchmarking," fixing model estimates to a specific time and set of survey conditions. We chose as our benchmark the six months in fiscal year (FY) 1987 in which the CPS had no supplements to confound estimates of the base CPS survey. We required that our model's estimates for "current methods" match average actual costs in these months across eight major accounting categories, and across subcategories for salaries, travel, and communications within them. To make these

comparisons, additional sets of LOTUS equations were required to map model-estimated costs into the categories and subcategories of the accounting system. Model parameters were then adjusted to obtain agreement at both the category and subcategory level. Agreement on salary costs generally was obtained by adjusting minutes per case until model caseloads and payroll salary costs were both satisfied.

A second form of validation was also employed. Since the numbers of required interviewers and supervisors are products of the model rather than inputs to it, these products also can be compared with actual figures by position, grade, and level for each major set of activities.

Benchmarking grounds estimates for current data collection methods in actual costs. Other procedures are necessary to project the costs of new methods for which current experiences are insufficient. We initially viewed CATI costs as better known than field interviewing costs. HTC activities were more easily observed, the CATI system automatically timed much of the work, and a research emphasis at the HTC mandated careful recording of costs.

Benchmarking CATI costs on FY 1987 expenditures could be highly misleading, however, since at that time the HTC was a relatively new and experimental facility. Initial small workloads made for operational inefficiencies, such as the use of supervisors for many interviewing and clerical tasks, which could easily be corrected in a fully established production facility. Other staff costs were spuriously low because personnel were concentrated in entry-level positions typical of the newly hired; and not all quality control procedures were yet in place, such as CATI reinterviewing and systematic monitoring. Cost estimates had to be "matured" by budgeting in all required activities, redistributing the work to more appropriate staff, and changing staff grades and ranks to those of a steady-state operation.

3.2 Model Presentation Methods

Methods of model presentation were improved. LOTUS 1-2-3 definitions, variables, and equations are convenient for model building and estimate preparation; but they are difficult for the uninitiated to read and comprehend. This became important to model development in two ways.

First, in an attempt to validate the model's code line-by-line, we called meetings of field work experts to review the definition and input value of each variable and the appropriateness of each equation. These meetings proved tedious, time-consuming, and relatively unproductive. All participants found it difficult to review the model piecemeal without understanding how these pieces fit together. Decisions to validate the model by accounting categories for a specific time period rather than by individual lines of code grew from these meetings; but some line-by-line review was still required.

Second, we encountered virtually the same problem in presenting early versions of the model to managers and decision-making bodies. They readily understood the purpose of the model, its general approach, and the estimates prepared; they could not follow many assumptions and the

relationship of one part of the model to another. The benefits of higher-level review could not be realized.

Our proposed solution is to present the model as a series of figures similar to flowcharts which clarify the relationships among its components. Figures 1, 2, and 3 illustrate this approach. To review specific scenarios, specific input or derived values can be shown for each box. Whether this approach sufficiently simplifies model presentation to expedite line-by-line validation and decision-maker review remains uncertain. We are still learning how to use these methods more effectively. However, graphic methods have facilitated our own discussions of the model and identified errors not apparent in the LOTUS 1-2-3 definitions and equations.

3.3 Sensitivity Analysis

The cost model's most tenuous assumptions are those on field interviewing time and mileage after current telephone interviews are transferred to CATI. The CPS field interviewers currently complete both personal visit and telephone visit interviews, conducting the latter from their homes. When the telephone interviews are transferred to CATI, the field interviewers will be left with a predominantly personal visit workload. Moreover, to maintain current interviewer workloads (and compensation), remaining field interviewing assignments must be revised to encompass larger geographic areas. The net effects on interviewer minutes per case, on miles traveled per case, and on other travel costs cannot be easily predicted in advance.

For the Two-Phase CPS, our best guess is that interviewer time and mileage will both increase by 25 percent per case over that currently required for field personal visit interviews. Mechanically, all adjustment factors (the A_{ij} in equations of Section 2.2) which apply to the interviewing time, mileage, and other travel costs of personal interviews in CATI eligible areas, were changed from their base values of 1.00 to 1.25.

To test the impact of this assumption on total estimated unit costs, a sensitivity analysis was conducted. The "best guess" input value of 25 percent was allowed to range up to 50 percent in 5 percentage point intervals using the DataTable feature of LOTUS. The impact on total data collection costs per case for the Two-Phase CPS is shown in Table 3. The simple linear increase in the linked input variables is reflected in the

Table 3 -- Estimated Data Collection Costs Per Case For The Two-Phase CPS By Assumed Percentage Increase In Field Interviewing Time and Mileage

Percentage Increase in Interviewing Time and Mileage	Collection Cost Per Field Case	Collection Cost Per Case (All Cases)
25 percent	\$34.70	\$20.55
30 percent	35.80	21.05
35 percent	36.90	21.60
40 percent	38.00	22.10
45 percent	39.10	22.65
50 percent	40.20	23.20

correspondingly linear effect on the output estimates. When more empirical evidence is available to assess differential consequences on interviewer time, mileage, and other costs, more sophisticated sensitivity analyses will be possible.

Sensitivity analyses have proved to have additional benefits. Component as well as total unit costs were examined for each selected value of the adjustment factors, and those which varied (or failed to vary) unexpectedly with the adjustment factors often suggested needed model changes or previously undetected errors in the equations.

4. Future Plans

Development of the CPS cost model was initially viewed as a short-term assignment which would terminate once the impact of CATI and CAPI data collection on the CPS was estimated. Instead, the CPS cost model has evolved into a continuing resource for planning and decision-making. The model's parameters (and structure) are frequently updated as new validating information is obtained, enhancements are made, and additional scenarios are considered. The model's value for budget preparation should grow as the proposed scenarios draw closer.

Several extensions and enhancements of the model are clearly needed. The model currently provides estimates only of data collection costs of the core CPS. Extension of the model to the CPS supplements is planned but faces problems of benchmarking data collection activities which change almost every month. First steps also are in progress to include data capture, editing, and data processing costs in the model. Estimates of total survey costs are now prepared by combining model estimates for data collection with cruder summary values for other activities, but more analytic approaches are being explored.

The model currently is based on average national figures, but subnational applications are being considered. A simplified copy of the model has been sent to each of the Census Bureau's twelve Regional Offices for trial use as a management tool. When national parameters are replaced by their regional values, the Regional Directors may find the model a useful way to compare projected with actual costs. We are also hopeful that regional use will yield regional parameters to enhance the national model. A composite model basing national estimates on a sum of validated regional parameters may eventually be possible.

In time, the model also must address the multi-survey nature of the Census Bureau's work. Both headquarters and Regional Office staff typically work on many different surveys each month; and the total volume of work affects the costs of each survey. A multi-survey model is required which seeks optimal staffing solutions across surveys and which can estimate the costs of any one survey taking the volume of work on other surveys into account. A multi-survey model

should be an aid both in planning and in managing surveys.

While these plans may seem highly ambitious, and clearly will require many years to bring to fruition, our experiences with survey cost modeling to date have been sufficiently promising to continue active development of this new tool and to recommend it to others.

Footnotes

- 1/ This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.
- 2/ Other Census Bureau staff members have contributed importantly to the model's development and validation. Dave Christopher played a critical role in devising methods of model validation and benchmarking; and Greg Russell provided many key estimates for the Two-Phase CPS. Other past and continuing contributors to the CPS cost model include: William O'Leary, John Lipp, Helen Montagliani, and Kathleen Creighton.

References

- [1] Bryant, Karen J. and Lynn Weidman. "Developing Cost Models for CATI Surveys." Proceedings of the 1987 American Statistical Association Section on Survey Methods. Washington: American Statistical Association, forthcoming.
- [2] McCarthy, William F., "OR/MS Applications in Computer Assisted Telephone Interviewing (CATI) Survey Research." Paper presented at the Joint National Meeting of the Institute of Management Sciences and the Operational Research Society of America, April 1988.
- [3] McCarthy, William F., Helen S. Montagliani, and Lawrence S. McGinn. "NCS/CATI Cost Analysis." CATI Research Report No. NCS-1 of the U.S. Bureau of the Census, Computer-Assisted Interviewing Central Planning Committee, CATI Research and Analysis Subcommittee, September 1988.
- [4] U.S. Bureau of the Census. "Computer-Assisted Interviewing -- Research, Development, and Implementation." A report of the Computer-Assisted Interviewing Central Planning Committee, October 1987.
- [5] U.S. Bureau of the Census. "Preliminary Evaluation of CATI Data Quality and Costs in the Current Population Survey." CATI Research Report No. CPS-2 of the Computer-Assisted Interviewing Central Planning Committee, CATI Research and Analysis Subcommittee, April 1988.
- [6] U.S. Bureau of the Census. "Final Report on Planning for a Post-1990 Redesign of the Current Population Survey." A report of the Oversight Committee of the Joint BLS-Census Committee on The Current Population Survey, April 1988.

FIGURE 1. CASE ALLOCATION FOR THE 5,500-CASE CATI SCENARIO

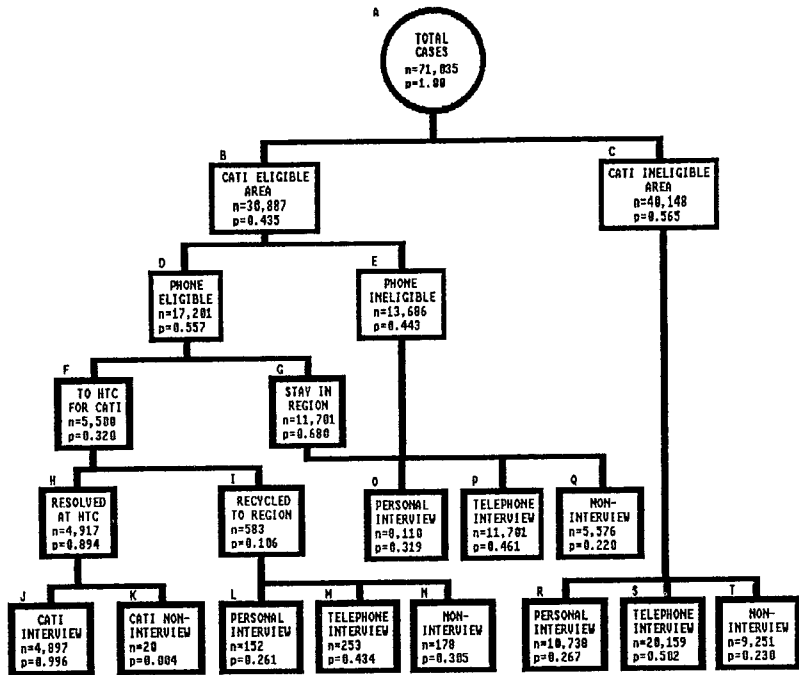


FIGURE 2. A PARTITIONING OF FIELD COSTS IN CATI INELIGIBLE AREAS

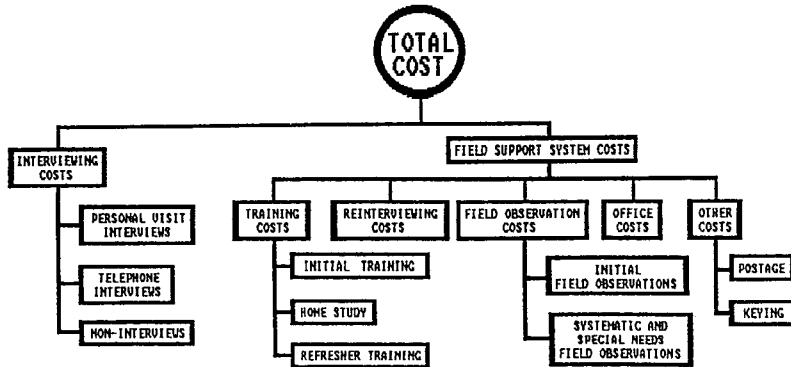


FIGURE 3. COST MODEL OF INITIAL INTERVIEWER TRAINING IN CATI INELIGIBLE AREAS

