WHAT WILL THE NEXT CENSUS COST:
The Use of Cost and Operational Modeling in the Examination of Alternative 2000 Census Designs

by Jay Keller and Carol Van Horn
U.S. Bureau of the Census

KEY WORDS: Cost model, census, 2000

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.

Executive Summary

The U.S. decennial census is a massive exercise that employs over one-half million persons, and costs over $2 million per hour during it peak operations. It is also complex and intricate, with a multi-year planning cycle, unique technology requirements, large preparatory tests, and a lengthy post-census publication schedule.
Computer models have been utilized at the Census Bureau since 1985 to examine data collection staffing requirements, productivity and scheduling. Current models have been refined to include all aspects of the census process. These utilize as many as 45,000 equations and relationships, and offer planners a sophisticated way to examine the cost and operational implications of changes to the design of the next census. Analyses to date include the impacts of content reduction, sampling or truncating major operations, matrix sampling, the use of administrative records, and a two-stage census.

Introduction

The U.S. decennial census is a massive exercise that employs over one-half million persons, and costs over $2 million per hour during it peak operations. It is also complex and intricate, with a multi-year planning cycle, unique technology requirements, large preparatory tests, and a lengthy post-census publication schedule.

Near the beginning of the planning cycle for the 1990 census, data collection planners and managers determined that the rapidity of shifting assumptions and changing scenarios for such a large-scale exercise called for an automated ability to assess the implications of potential changes for staffing plans and proposed budgets. Subsequent to this initialization of an automated cost modeling program for data collection activities, models were also developed for data capture and processing concerns.

The success of these efforts led to a program, begun in 1989, to use computer models to examine the entirety of the decennial census process. This examination would be a tool used in assessing proposed design changes for the 2000 census, and planners have considered cost as a key criterion in the assessment of alternative designs.

This paper reviews the development of the computer modeling effort as a planning and operational tool for the decennial census, examines results to date in the assessment of alternatives considered for the 2000 census, and looks ahead to the next steps for the decennial census modeling program.

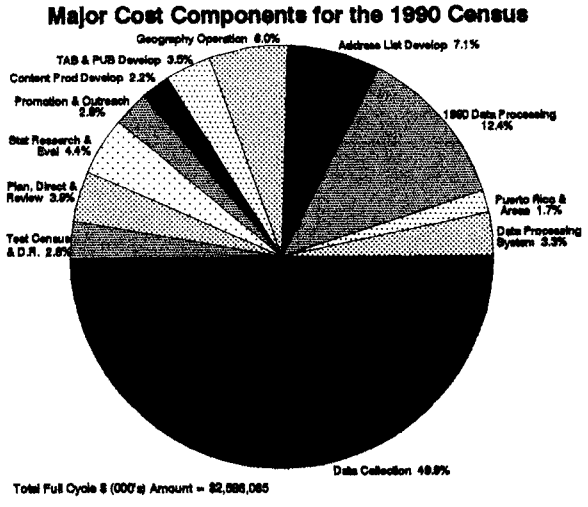
Part One: What Did the 1990 Census Cost?

The Census Bureau will not know until after September 30, 1993 what the 1990

census actually cost, since some activities are continuing through the end of this fiscal year. Actual results through fiscal year 1992, and estimates for 1993, show a total cost slightly less than $2.6 billion.

This total cost can be viewed in many components (see chart 1), but by far the greatest piece is for 1990 data collection. This includes the leasing of space for 449 district offices, the salaries of full-time managers and assistant managers who ran the offices, the support staff who recruited, hired, and paid the temporary employees who staffed the data collection operations, and the logistical requirements of establishing telephone lines, printing and shipping training and operational manuals and forms, installing computer hardware and software, and supplying an enormous work force.

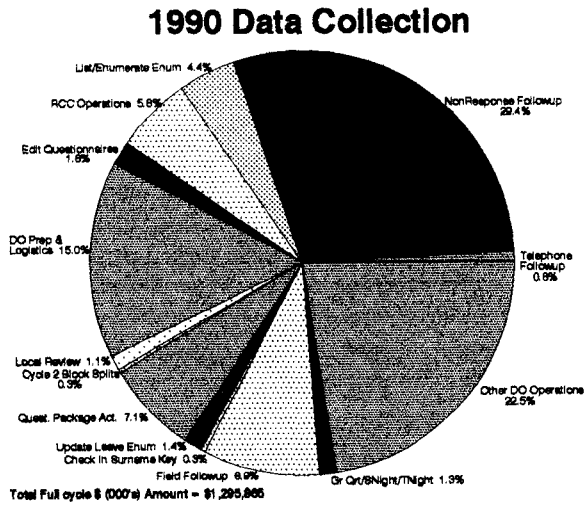
CHART 1



Operationally the data collection effort consisted of several dozen field and supporting office operations, which refined address lists, followed up on census nonrespondents, enumerated special places and group quarters, followed up on edit failures, and prepared and controlled assignments for the enumerators who carried out the operations.

All of this took about half of the $2.6 billion budget, with the single biggest operational share going to nonresponse followup, which took roughly 30% of the data collection budget (see chart 2). This operation, the work load for which is directly controlled by the nation's total housing unit count and the rate of mail response to the census questionnaire, drives the sizing of the district offices and management and support staff.

CHART 2



Data capture and processing, and data collection, make up less than two-thirds of total census costs, and it's necessary in a review of the 1990 census, and for future cost projections, to consider the other pieces of the census cost pie. Some of the remaining cost is in the address list development operations conducted prior to the census year, especially the national prelisting and precanvass operations--conducted on foot by census enumerators--which build and improve upon the address list. There are also geographic activities, and the census work in Puerto Rico and outlying areas.

But a sizeable portion of census costs come through our headquarters and planning activities. Formal planning for the 1990

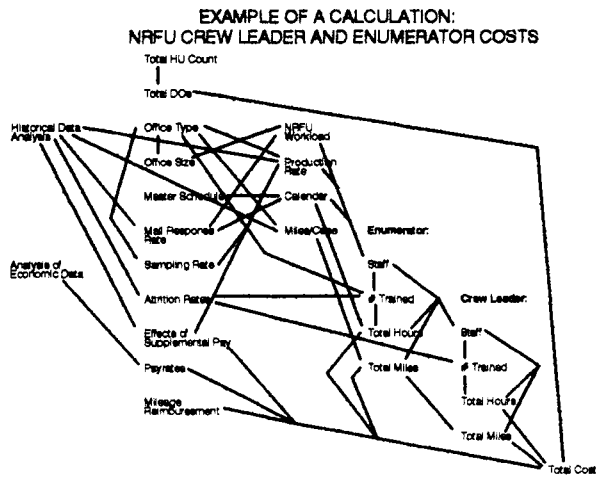
census began in 1984, and major tests were conducted in 1985, 1986, and 1987. These each utilized between one and three full-scale district offices to test methods and innovations, including new computer hardware and software, for the next census. In 1988 a dress rehearsal of the 1990 effort was tested in the states of Missouri and Washington. Significant expenditures throughout the decade also went to statistical research and the evaluation of results, to content planning, to promotion and outreach efforts, and to the development of questionnaire content and format. And work continues in tabulation and publication, which before the census required a long lead time for its developmental work.

Considerations of the cost of the next census must look carefully at each of these items, and not just expectations for improved or new data collection and capture methodologies.

Part Two: The Modeling Program for the 1990 Census

District office data collection modeling began in 1985--for a 1986 census test in Los Angeles and Meridian, Mississippi--using personal computers and Lotus 1-2-3 software. The models depicted each data collection operation, staffed in the field by enumerators or in the office by clerks or keyers. A multitude of variables considered each individual operational workload--based on the total size of the district office, and the functional relationships among each operation--as well as production rates, schedules, supervisory ratios, turnover, payrates, mileage reimbursements, second and third shift work and pay differentials, and the use of bonus pay for superior productivity. As can be seen in a view of the flow of calculations to determine the cost of a single operation, the components of data collection cost determination are many, and their interrelationships somewhat complex (see chart 3).

CHART 3



Because of the opportunities resulting from the test census structure, we were able to refine the data collection model and projections based on actual results from the field. By the time of the 1987 district office test in North Dakota, the model was available to our district office manager to update as actual work load information became available, and this recast staffing requirements and total budgets for hours and miles available to accomplish the job.

By 1988 the modeled interrelationships of operations allowed macro level results as well as results for individual operations. This meant that models could also derive square footage requirements for offices, staggered telephone line installation and usage requirements, total recruiting goals, and various miscellaneous items involved in the establishment and functioning of a district office, as well as producing training and production staffing levels for individual operations. The models had also become key tools for headquarters planners developing training and procedural manuals with staffing, productivity, and scheduling expectations, and program coordinators concerned with the day to day data collection function.

Also by 1988 the models started to receive use for nationwide efforts. Address list development efforts--the national prelist and precanvass--in 1988 and 1989, which cost about $182 million, had their data collection budgets fully derived from cost models. And similar to their use in district office tests, field managers of these operations used the models interactively for staffing and budget projections. Refined over a several year period, the models were extremely accurate.

For fiscal year 1990 models were the basis of the Field Division's nearly $1 billion budget, including the full costs of 449 district offices and 13 regional census centers which managed the district office functions. Again in the hands of field managers, models provided the budget side of daily budget versus actual reports which informed managers of progress at the local, regional, and headquarters levels. Management reporting was also facilitated by the use of daily payroll forms, keyed into a payroll data base and a cost and progress system, which provided the actual side of the comparison of hours, miles, dollars, total cases complete, and production rates by operation. This was only possible because of a painstaking and long-term revision to the categories in which field costs were collected, and through the use of more than 90 separate operation codes.

Used interactively in this way, cost models did more than project requirements, they became a tool in reducing costs and maintaining a schedule. One series of budget versus actual reports, at the enumerator level, allowed managers to compare daily productivity among employees and to budgeted standards. This facilitated the rapid dismissal and replacement of enumerators--necessary in staff-intensive operations like nonresponse followup--and removed low producers from the payroll. The interactive use of the modeling tool and progress reports also allowed managers to identify potential savings in certain operations which they then redirected to potential problem areas. By the end of the census year, the completion of all but one of the district office operations, and the staggered closing of 449 offices, the field division had experienced actual costs that were within two-tenths of one percent of its budgeted expectations.

Part Three: Modeling 2000 Census Alternative Designs

We continue to use Lotus 1-2-3 for a full Year 2000 cost model, which we call The Year 2000 Input/Output model. The more generic title reflects the more general uses of the tool: It projects staffing requirements and other resource levels, allows sensitivity studies for a variety of operational issues, and allows the production of ranges of information and various scenarios for each alternative design it has examined.

Using Release 3 of the spreadsheet software we have been able to make use of greater memory efficiencies, and also a three dimensional spreadsheet capability. The prototype of this model ran on a 286 personal computer and took in excess of 30 minutes to calculate changes within a set of variables. The model, which currently has in excess of 44,000 equations, relationships, and references, and utilizes about 4.5 megabytes of random memory, now runs on a 486 PC with enhanced memory, and calculates in less than one minute.

The use of spreadsheet software resulted from a recommendation by a consultant the Census Bureau engaged in the late 1980's, as well as from the agency's interdivisional familiarity with this tool for financial analysis, even within the administrative budget process. Other software, including statistical software, is utilized for the tangential studies that support key model assumptions.

The model is complicated. Individual functions range from simple to complex, but the real intricacies are in the sheer size of the

package, and the interactions and interrelationships contained within it. This results from the size of the census itself, which in a full cycle involves hundreds of projects and operations, each with a great number of variables. It results also from our goals of using modeling results operationally and administratively. These requirements add the need for a bottom-up derived level of detail which is not present when the goal is simply to look at a macro level of impact.

But it is also true that simpler measures, such as a cost per case approach to looking at census costs, do not work well, and that real functions are too complex for simple models. Costs per case, for example, are often not equal depending upon methodology or area. A clear example lies in the issue of sampling for the nonresponse followup operation, an option for which there is great expectation for potential cost savings. This savings, however, cannot be derived by measuring the cost per case of nonresponse followup from the 1990 census, and multiplying this by the number of housing units which would be excluded in a sampling scenario. First, the actual measurement of the cost per case is unclear. Is this only the cost of the physical enumeration, the hours and miles incurred by the employees? Does it include some percentage of the full-time management of the district office? The space? What about the next tier of management, the regional census centers utilized in 1990? Should it include some component of the headquarters planning and coordinating activities?

Even if properly measured, a cost per case provides insufficient information for the analysis of changes to designs or operational plans. In the nonresponse followup sampling scenario, cases that are sampled might be further apart from each other than cases in a complete count. This would increase travel requirements, and act to decrease overall production rates. Thus, the cost per unit would probably be greater, and the cost per unit analysis from complete count data would be a poor measure of the overall impact.

Indirectly, a cost per case analysis misses other potential impacts, such as changes in managerial staffs or strategies. Perhaps, for example, fewer field offices are needed to accompany a smaller field staff in a sampling scenario. In general, all of these issues argue for models of increasing complexity as the complexity of the program or organization increases, as well as a sophisticated charting of the interrelationships across operations and cost components.

The increased level of detail also assists with the obvious, inherent problem of domain of applicability. Our experiences, and operational and cost data, are from censuses of particular designs. In assessing alternatives to different designs our level of certainty about the values of key variables diminishes.

There are several ways to deal with this problem. One is through the use of tests, either small directed experiments or full scale district office tests, which provide information about the impacts of designs changes. Model variables can be revised based on the new information. In the absence of field testing, variable values can be extrapolated in other ways. Not infrequently, managerial assessments are used to determine ranges around key variables due to proposed changes.

Experience has shown that these methods, coupled with detailed modeling that allows the full consideration of small and large variations, yield accurate results. For the 1990 census, operational work flows for questionnaire editing, followup, and control were adjusted to a design untested before the actual census, but models were able to accurately project the cost and staffing implications.

Innovations that had significant impacts on one of the key census data collection variables--the productivity rate for the nonresponse followup operation--were also not fully tested before 1990 due to late operational

improvements. These included a system of differential pay rates, a new scale of enumerator incentive payments, and the impacts of the managerial reporting system. The impacts of similar programs, however, had been measured--sometimes separately, sometimes in combination--at other times in the decade, allowing for an extrapolation of the value of the variable. The estimates for this crucial item, which directly impacts the massive staffing level for the nonresponse operation, were extremely close to what was subsequently realized in the field.

Results of Alternative Year 2000 Design Modeling

The census design scenarios examined to date in the program for their cost implications include those aimed at reducing content and simplifying the questionnaire, spreading out the work load of the census data collection effort, and reducing the work load directly through sampling or truncation.

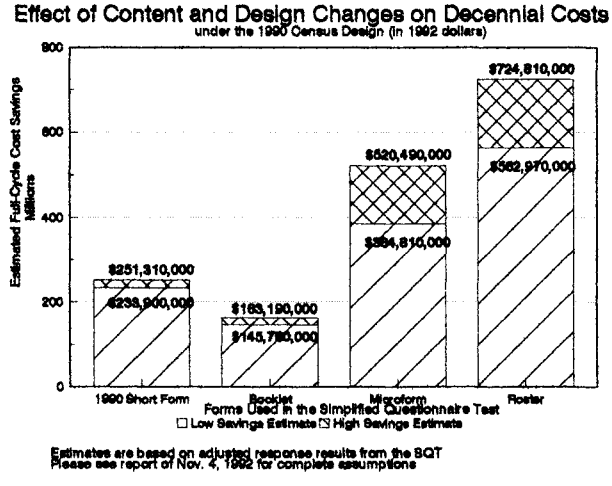
One alternative is to reduce the size of the sample receiving the long questionnaire to zero. This would facilitate content reduction goals in a census, or provide the first phase of a two-stage census process, or provide the census year component of a continuous measurement methodology.

There are several impacts of this on census costs. Because the mail response rate to a short form is generally higher than response to the long form, the average mail response rate would increase slightly. This works directly to reduce nonresponse followup and its associated costs. Next, average interviewing time decreases with the removal of the long form from the enumerator's assignment, again reducing staffing needs. Fewer edit failures reduce workloads in edit and telephone followup operations, and increase productivity. Questionnaire printing and mailing costs decline due to the use of only the short form.

Modeling results project that a 1990 census using a short form only approach would have cost from $233,900,000 to $251,310,000 less than 1990 full cycle costs, expressed in current year dollars ("Analysis of the Full-Cycle Cost Implications of Reducing Content and Altering Questionnaire Design and Implementation as Conceptualized in the Simplified Questionnaire Test," Susan Love, U.S. Census Bureau, 1992.)

Successive reductions in content and the elimination of questions from the questionnaires offer the potential for corresponding increases in savings in the census year (though if new efforts were added to obtain the lost data in other ways, this could provide offsets against savings). The Census Bureau has tested several new instruments in this regard. One of these, a "microform" tested in 1992, which contains 4 of the 7 census 100% population items, arrayed on both sides of a single sheet, and no housing questions, would yield a higher mail response rate, further reductions in printing cost, and still faster interviewing time. Potential savings here are estimated in a range from $384,810,000 to $520,490,000. (Ibid.) At the extreme of content reduction strategies, a roster requesting only name and date of birth, printed on cardstock, might yield savings as high as $724,810,000 (see chart 4).

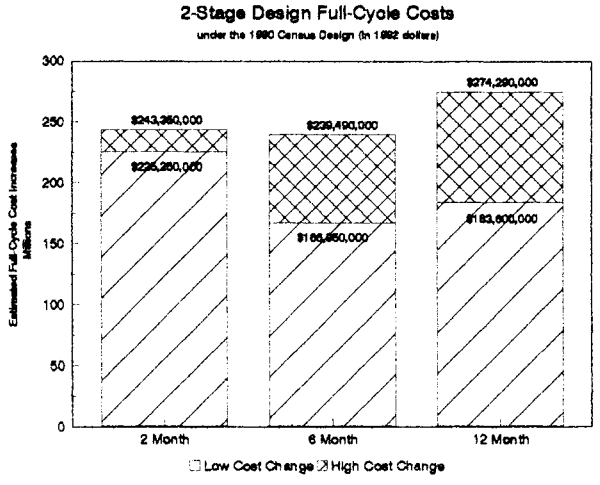
CHART 4



Other designs seek to spread out the workload of the census. One benefit is a reduction in peak census staffing. In a two-stage census design the total time period of the field data collection activity would be increased, with only 100% data collected initially, and sample data collected in a second stage of operations. The information gained from the first stage of the activity may allow a decrease in the size of the census sample.

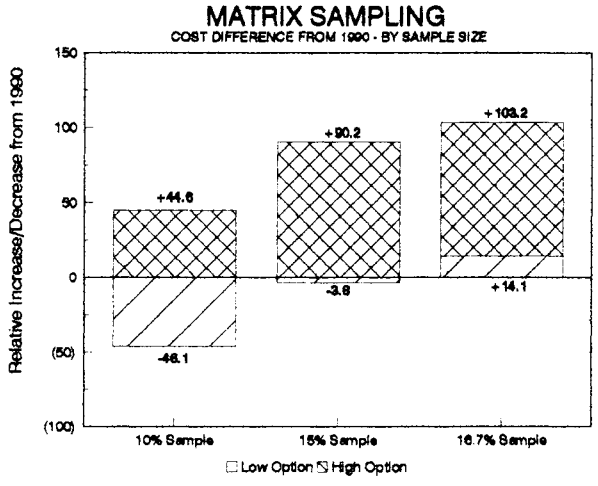
Regardless of the specific time period between stages (this was examined with a two month minimum, and a twelve month maximum time interval), total census costs were shown to increase. This is primarily due to the impacts on district offices, now open longer, with extended employment for managerial and support staffs. With sampling rates in a range from 1990 rates to a 12% reduction in 1990 rates, and with a range of time intervals between stages, cost increases to the census may be from $225,250,000 to $274,290,000 in 1992 dollars, and compared to the 1990 census. Additional studies showed that to avoid increases in total census costs due to the use of this methodology, the sampling rate must be half or less the rate utilized in 1990.

CHART 5



Matrix sampling--the use of multiple versions of the long form, each with potentially fewer questions to reduce individual respondent burden--is estimated to also increase census costs. Increased training would be necessary for enumerators using multiple forms for enumeration. Tabulation and publication costs would increase. Overall productivity in enumeration and edit control operations would decline slightly (though the decline in nonresponse followup productivity may be partially offset by decreased long form interview time). Printing and postage costs would increase as the overall sampling rate increases. However, the net impact is slight, increasing census costs in a range from 1% to 5%. ("Matrix Sampling--Cost Information and Analysis," by Ed Kobilarcik, U.S. Bureau of the Census, May 1993)

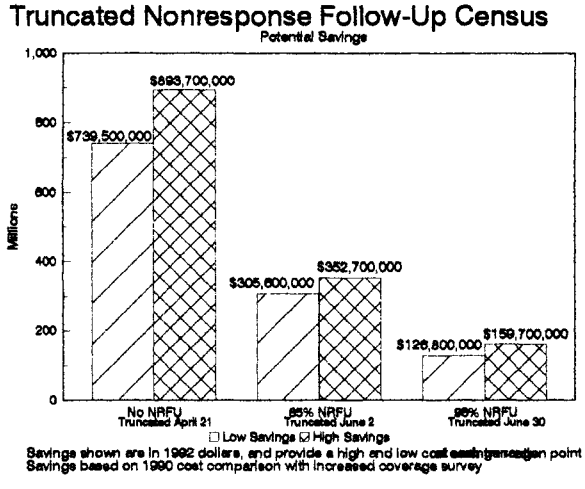
CHART 6



The truncation of the nonresponse followup operation is an alternative that has been examined at three specific truncation points: with 98% of the nonresponse followup work load completed, with 65% completed, and with a full elimination of the operation. Such an alternative would reduce staffing requirements and census costs, but require increased estimation of the data no longer collected. Small reductions in work

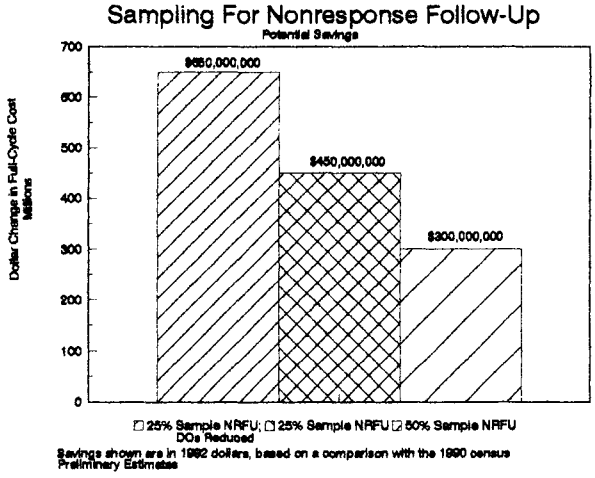
loads here would result in savings from hours of production work by nonresponse followup enumerators. Increasingly larger reductions could reduce supervisory requirements, support staff requirements (for recruiting and payrolling), and office space and equipment. The extreme reduction of 100% of the operation suggests reducing the number of district offices necessary to support field activities. The minor truncation of the operation may allow offices to close earlier, and the discontinuation of some operations subsequent to nonresponse followup, including the followup check of vacant and deleted housing units. Savings here are estimated to range from $126,800,000 to $159,700,000. A truncation which completes 65% of the work load may save in a range from $305,600,000 to $352,700,000. Finally, a complete truncation of the operation and its correlated impacts may save from $739,500,000 to $893,700,000. Information on the potential cost savings of a truncated census may be found in "Analysis of Potential Cost Reductions for the Truncated Nonresponse Follow-Up Census."

CHART 7



Another vehicle for reducing large nonresponse followup costs is to sample the operation. Here, sampling might be devised on either a block or case basis, and may range from very high sampling rates to those as low as 25%. While productivity will drop and miles driven per case by enumerators will increase, overall the methodology would reduce staffing requirements, and at lower sampling rates, managerial and facility overheads. Total savings at a 50% sampling rate might exceed $300 million. At 25%, they could approach $650 million.

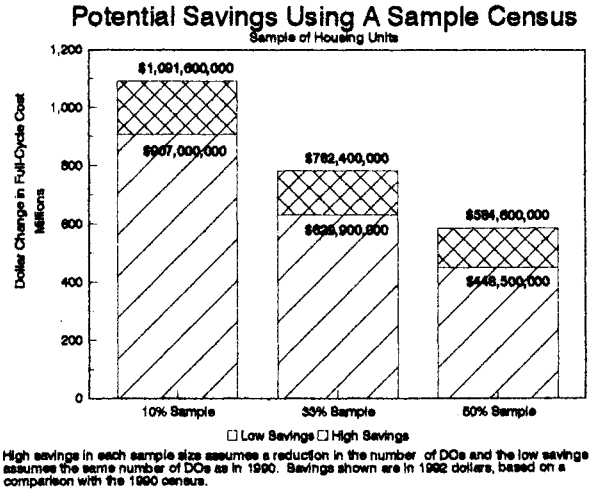
CHART 8



Finally, sampling for the total census count, at the initial mailout stage, is also a possibility, though we believe that conducting the entire census on a sample basis is unconstitutional. Reapportionment and redistricting data, and age and race/ethnic data defined by law to meet the requirements of enforcing the Voting Rights Act, would all be based entirely on a sample in such a design.

Scenarios include block and case sampling methodologies, and sample sizes ranging from 10% to 50%. Such a methodology would have extreme impacts on all aspects of the data collection and processing activities, and a range of potential savings from $449,000,000 to $1,092,000,000.

CHART 9



Conclusion

Subsequent work in the area of operational and cost modeling of the Year 2000 Census will include close examinations of censuses preceding 1990 in an effort to understand fully the comparison of costs across censuses and the impacts of different methodologies, levels of planning, and quality control on census activities. Future design costing will also look at continuous measurement and administrative records censuses. The finalizing of the 1995 test design will also lead to full operational modeling in support of that test, including the derivation of staffing and facility requirements, and the link between authorized expenditures in the data collection and data capture activities and budget vs. actual reporting.

The results of the 1995 test, including its productivity and work load measures, will inform the modeling of the final selected design for the next census, and continue to facilitate the planning effort. The operational interrelationships, schedules, and budget requirements depicted in the automated model will be linked with internal managerial information systems. This was partially done in the 1990 census and will be fully utilized across all aspects of the census effort in this decade, allowing improved efficiency, increasing the control of costs, and building managerial effectiveness.