Continuous Measurement Alternatives to Census Data Collection

Keith F. Rust[1]
Westat Inc., 1650 Research Boulevard, Rockville, Maryland 20850

1. Introduction

Concerns of the United States Congress, the Commerce Department, and the U.S. Bureau of the Census with the cost and quality of recent censuses (especially 1990) are having a substantial impact on the planning process for the year 2000 Census. Two particular goals have been identified for the procedures in 2000. The first is to conduct the census at a lower total cost than in 1990, and the second is to reduce the differential undercount in the census enumeration. In recent censuses the number of persons missed has been low by historical and international standards, but those persons not enumerated by the census have been concentrated among certain demographic subgroups in the population. The aim is to decrease significantly the level of undernumeration in these traditionally undercounted groups, without decreasing the overall level of census enumeration.

Developments for the Census in 2000 involve research into activities aimed at meeting these two aim simultaneously. These include the development of integrated coverage measurement procedures, sampling for nonresponse follow-up, the use of administrative records, the development of special methods for use in specific geographic areas, and changing the date of the census. The Census Bureau's plans and thinking are reviewed in the report "Counting People in the Information Age" of the National Research Council (Committee on National Statistics, 1994).

A natural consequence of attempting to reduce both cost and differential coverage simultaneously is that efforts be made to find ways to simplify the process of enumerating the population. Some advocates of census reform have questioned the collection of detailed socio-demographic data as part of the decennial census. Since 1960, this additional information has been gathered by distributing a census "long form" to a national sample of households. One current argument is that the accuracy of the decennial population figures would be improved if long-form data collection were eliminated, reduced, or displaced in time from the effort to enumerate the population.

Questioning the inclusion of these data collection activities as part of the decennial enumeration operations leads directly to the broader question of determining the most appropriate manner for fulfilling small area data needs throughout the decade. With this in mind, one activity that the Census Bureau is undertaking that has generated considerable interest is the development and evaluation of a proposal for a collection system for small area data that is conducted throughout the decade, rather than concurrently with the decennial census. This plan, referred to as continuous measurement, is the subject of this paper.

Other approaches to reducing data collection in connection with the census have been suggested. One proposal is that some of the data that was gathered in the 1990 census could in future be collected through alternate methods, such as sample surveys or tabulations of administrative records, and made available for use in a more timely manner (see, e.g., Sawyer, 1993, and U. S. House of Representatives, 1993).

The next section of the paper reviews the motivation and history of interest in a continuous measurement program. The Census Bureau's current plans for testing and evaluating continuous measurement are then discussed. The remainder of the paper identifies some of the key issues that must be resolved in undertaking an evaluation of the proposed continuous measurement program. A more detailed discussion of this topic is contained in Chapter 6 of "Counting People in the Information Age" (Committee on National Statistics, 1994).

2. The Proposal for Continuous Measurement Program

Sampling was first used to gather additional information (or content) for the Population Census in 1940; in preceding decades, all questions were asked of all households in the census (Goldfield, 1992). In 1990, as in other recent censuses, the long form contained questions about occupation, income, journey to work, ethnicity, and housing. The long form is completed by a sample of census respondents; in 1990, the national sampling rate was one-sixth of all households, although the fraction was as large as one-half in some small units of local government.

---

[1]The author was a member of the Panel to Evaluate Alternative Census Methods, Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, National Research Council. This presentation has been informed by discussions and research that have occurred as part of the Panel's activities, and in particular the author is indebted to panel members and staff Norman Bradburn, Bruce Petrie, Ed Schillmoeller, and Duane Steffey for their contributions to this presentation. Presentations by, and discussions with, Census Bureau staff, and in particular Charles Alexander, have been invaluable in providing information and perspective on the planning for Continuous Measurement. The views expressed here do not necessarily reflect the opinions or recommendations of the Panel itself, nor any of its individual members other than the author. However, they are in accord with the recommendations included in the final report of the Panel, "Counting People in the Information Age", published in September 1994.

The long form includes all content that is contained in the short form completed by all other households. The 1990 long form contained a total of 33 questions for each household member and 26 questions about the housing unit. For the short form, the respective numbers were 7 questions per household member and 7 housing questions. Thus, the long form requires considerably more information from a given household than does the short form.

This paper considers an approach to reducing the long-form burden at the time of the census, that represents a fundamental change in methodology for obtaining data of the type traditionally collected on the decennial census long form. This method, referred to as continuous measurement, proposes to spread the collection of sample data for small areas and subpopulations (or "small-area sample data") across the decade. Continuous measurement has implications that extend far beyond issues of census cost and accuracy. Careful evaluation and widespread consideration of these implications will be needed to clarify the merits of this proposal.

There have been proposals in the past that collection of small-area sample data, such as that typically provided by the census, be conducted throughout the decade, rather than once every ten years. Kish (1981, 1990), Horvitz (1986), and Herriott et al. (1989) have proposed a variety of data collection schemes that involve this key concept of extending data collection in a more or less continuous fashion. As a part of its planning activities for the 2000 Census, the Census Bureau has included an evaluation of this type of process, and this evaluation is currently ongoing. In its Design Alternative Recommendation #14, released in 1993, the Bureau indicated a commitment to investigating fully the feasibility of continuous measurement as part of the 2000 Census development process. Recently, Alexander (1993) has proposed a way in which a continuous measurement program might be instituted in conjunction with the 2000 Census.

These proposals for continuous measurement share two main features: (1) virtually continuous data collection operations, instead of starting and stopping every 10 years, with ensuing benefits for data quality through the maintenance of a permanent enumeration staff, and improvement through continuous experience; and (2) more current census data throughout the decade (except perhaps for the smallest geographic units, for which updates of census data might be based on a ten-year moving average of sample data).

A distinguishing feature of the plan currently being developed and evaluated by the Census Bureau is that it calls for continuous measurement to be conducted in connection with a complete enumeration of the population at one point every decade, whereas most earlier proposals would replace the traditional census completely with a continuous measurement program. The present proposal assumes that a decade enumeration is required in order to meet constitutional requirements. This view is supported by a legal review prepared by the Congressional Research Service (Lee, 1993) and subsequent work by the Panel on Census Requirements (Committee on National Statistics, 1993). Thus, the purpose for proposing continuous measurement in these circumstances is two-fold: (1) to reduce the cost and burden of the decade enumeration by removing the need to collect small-area sample data as part of the decennial census; and (2) to improve the frequency, timeliness, and quality of small-area sample data.

A related consideration in judging the merits of continuous measurement is the relationship between the decennial enumeration and statutory requirements for information. A recent review (U.S. Bureau of the Census, 1994) found that legislative mandates exist for most of the items collected on the 1990 census long form, but the statutes do not specify that the data must be collected in the decennial census. The key question is whether the decennial census is the most appropriate vehicle for collecting this information. Considerable research and development must occur to answer that question and thus determine the extent to which continuous measurement might replace long-form data collection, as opposed to supplementing it.

The ongoing thorough review of census requirements, costs, and methods presents an opportunity to undertake a full evaluation of the viability and desirability of instituting a permanent continuous data collection program to obtain traditional census data on population and housing characteristics. The efforts by Alexander and his colleagues to develop a prototype for continuous measurement appear to provide a very promising start to this process. An especially encouraging aspect of the Bureau's plans is that a project team has been established and has begun to carry out work on a program of evaluation for continuous measurement. Considered judgments about the merits and drawbacks of continuous measurement can only be made following the collection of considerable detailed information as to exactly how it will operate. The Bureau's initiation of a project team, along with a development plan, give cause to hope that adequate concrete information will become available through a program of research and development.

Much of the interest and discussion surrounding continuous measurement to date has concerned its cost when fully implemented. Based on the evidence seen to date it seems unlikely that continuous measurement would provide a less costly alternative to the traditional long form (see Section 5 for further discussion of cost issues). What continuous measurement is most likely to offer is improved quality (in the broadest sense) for the sample data (including currency throughout the decade), and perhaps for the decade enumeration as well. Benefits for the enumeration might result from the removal of the requirements to collect and tabulate sample data as part of the decade census operation.

3. Overview of Current Plans

Alexander (1993) identified three goals for the research on continuous measurement:

- Determine the basic prototype design for data collection and estimation;

- Estimate the cost of the operation, or at least give useful upper and lower bounds; and

- Make general statements about the quality and utility of the data (including coverage and content) from the continuous measurement system compared to other alternative systems.

Alexander goes on to identify some decisions that were taken initially in the research process, concerning the form of a continuous measurement design.

1. The continuous measurement prototype will include a complete year-zero (i.e., end-decade) enumeration for reapportionment and redistricting, rather than a "rolling enumeration".

2. The frame for intercensal samples will be the Master Address File (MAF).

3. The prototype assumes implementation in time to replace sample data for the 2000 Census, rather than waiting for 2010. The development plan is based on the assumption that a decision as to whether to replace sample data from the 2000 Census with a continuous measurement operation will be made in 1997.

4. The Continuous Measurement prototype must produce data for most 1990 "long form" characteristics for small areas, meaning tracts/Block-Numbering Areas and block groups, with more or less the same reliability requirements as the 1990 "long form" sample.

5. The prototype should assume direct sample-based estimates for small areas, rather than relying on model-based "indirect" or "synthetic" estimates or administrative records.

6. The basic small area (tract/BNA or block group) estimates should be rolling accumulations (moving averages) of five years of data. Current plans call for a three year moving average for 1999-2001 (with a corresponding increase in the monthly sample size).

7. Data collection will be spread evenly across the year and across the nation.

8. The survey will be a separate survey rather than an expansion of any current Federal household survey.

9. The design should include a combination of mail, telephone, and personal-visit interviews.

The Census Bureau has developed a schedule for implementing a research program into the feasibility of conducting a continuous measurement operation. This plan is shown in Exhibit 1, which is extracted from Alexander (1994). It indicates an expanding effort at developing a full system of continuous measurement over the next six years. This plan is characterized by three features: (1) a steadily increasing level of resources over time, from a relatively modest research effort in 1994 to the full system (as currently envisaged) in 1999; (2) a series of decision points at which the results of the research to date and other developments are evaluated, and a decision made whether to proceed with plans for a continuous measurement operation in place of "long form" data collection in connection with the 2000 Census; (3) parallel efforts at developing data collection capabilities, the estimation system, reliable cost estimates, and user needs, throughout the development cycle.

Staff at the Census Bureau have been pursuing research and development of a prototype system for continuous data collection. Proposals for the prototype have undergone several revisions, particularly with regard to such characteristics as sample size and date of initiation. The current proposal (Alexander, 1994) involves a random-digit dialing (RDD) survey starting in November 1994 (FY '95) at three to four geographic sites, totaling approximately 2,000 households per month.

The Census Bureau plans to use the data from the survey to produce six- or nine-month cumulations, in the form of data files for prospective users, in mid-1995. These data files would have the same format as that proposed for the mature continuous measurement program, in order to provide users with a sense of what estimates and data products would be available.

In October 1995, the RDD survey would be replaced by an address-list-based test in four sites involving a total of about 4,000 households per month. The sample size corresponds to the level of sampling that these sites would receive under the national system as currently proposed. Experience from actual field operations would be used to refine cost estimates.

The Census Bureau has begun to work with census data users at some federal agencies to study implications of continuous measurement. In particular, the Bureau has established contact with officials at the Department of Transportation and at the Department of Housing and Urban Development. Data users in state and local governments and in the private sector will also be consulted; such efforts will intensify when prototype data products become available.

Three key decision points for the continuous measurement program are:

1. October 1995: The Census Bureau could stop before increasing to the proposed level of effort at this point -- for example, if cost estimates for a continuous measurement program increase substantially -- but the program does not call for a large investment at this stage.

2. October 1996: Extensive data collection would begin so as to provide estimates for congressional districts, and to work on remaining problems. Users will have had time to consider the demonstration files from the FY '95 program and some of the FY '96 results. A conditional decision would be made to drop the long form from the 2000 census if the FY '97 program is successful.

3. September 1997: A final decision would be made about whether to retain the long form for the 2000 census, replacing it with the continuous measurement program. One possible complication in reaching this decision point is that, in early 1997, the Census Bureau must provide, for congressional review and approval, a list of topics to be included in the 2000 census. It is unclear whether the means of collecting data on these topics must also be determined by early 1997.

If continuous measurement were not implemented for the 2000 census, then the Census Bureau would try to maintain program activities, such as updating the Master Address File, using current survey interviewers. The Master Address File could serve as a frame for household surveys, such as the Current Population Survey. Note that a decision to use a long form as part of the 2000 Census does not necessarily imply that continuous measurement should not be implemented during the next decade.

If continuous measurement is implemented in place of a long form in 2000, then the sample size of the monthly survey would increase from approximately 80,000 households in 1997 to 100,000 in 1998 and, finally, to 325,000 in 1999. In 1997, estimates would be available for states and other large areas. A systematic (and geographically sequenced) sample of households within census blocks would begin in 1999.

The Census Bureau has now established a project team to carry out the research into the practical issues in developing a system of estimates obtained from a continuous measurement program. This group will be responsible for the development of the RDD telephone survey at the four test sites, and subsequent development into an address-list-based test. The group will examine various aspects of questionnaire and content issues, working with interested parties and advisory panels, evaluating cost components, developing operational systems, developing the survey design and estimation procedure, and evaluating the impact on current household survey programs throughout the federal statistical system.

The breadth that is encompassed in this early thinking that has taken place in developing the plans to examine the desirability and feasibility of a continuous measurement system is impressive. It is important to recognize the need to develop the various aspects of the research effort in parallel. This is essential if an effective continuous measurement system is to be developed. The evaluation of each of the aspects of the system depends upon the research efforts in the other areas. To study whether user needs will be met effectively by continuous measurement, it is important that simulated products be provided to users at an early stage in the investigation as proposed by the Bureau, and that user comments be solicited to guide the process. These user responses may well impact the requirements for data collection and estimation. On the other hand, preliminary data collection and estimation procedures must be in place to develop the simulated products. Thus, all features of the system must be developed in a synchronized fashion.

While this might appear obvious at this point, it will be necessary to maintain a strong commitment to this principle of a synchronized development over time. It is easy to envisage that some aspects of the research will be easier to develop, have a clearer path, and be under the Census Bureau's control to a greater extent than others. One can imagine that a situation could easily develop where research into the data collection methods is progressing rapidly, reaching an advanced stage with many decisions finalized, while essential research into user needs is lagging and hence not informing appropriately the decisions on data collection methods. Vigilant efforts will be needed to ensure that uniform and timely progress is made on all fronts.

The succeeding sections address in more detail various features of the continuous measurement operation that is envisaged. First, a summary of the key features proposed by the Census Bureau, as outlined in Alexander (1993) is provided.

The program as envisaged by the Census Bureau combines three main components: (1) continuous updating of a national master address list; (2) a large periodic sample survey to collect intercensal data, using the master address list as a frame; and (3) an "Integrated Estimates Program", which produces estimates from the periodic sample survey, using other data sources such as the decennial census, the master address list, and administrative records to enhance the estimates. The data and estimates from the continuous measurement program will in turn be used to enhance estimates from other national household surveys and the demographic estimates program.

The Intercensal Long Form (ILF) will sample about 250,000 addresses (as currently hypothesized) nationally

each month, drawn from all geographic areas. Different addresses will be included each month. The initial sample will be mailed a data collection form. A subsample (possibly all) of the mail non-returns will be followed up by telephone where possible. A further subsample of those who cannot be contacted successfully by telephone (for whatever reason) will be followed up in person.

Annual average estimates will be produced for large geographic areas where the population exceeds 250,000; such as states, large MSAs, and groups of counties. For small areas, such as tracts and block groups, a moving 5-year average will be produced annually, using data from the previous five years.

The survey frame of mailing addresses will be updated regularly using Postal Service mail delivery lists, and possibly lists from local governments. The list will be updated quarterly. City-style addresses will be geocoded to the block level. For areas with non-city-style addresses (rural delivery routes, post office boxes, general delivery, physical description only), additional efforts will be needed (see Alexander, 1993). The extent to which these efforts are needed depends upon the extent to which city-style addresses are prevalent on the Master Address File (MAF) in 1997. The feasibility of the procedures for non city-style addresses, and their extent, are important in establishing the viability of a monthly national ILF survey.

Once the data have been collected, survey weights will be applied. Initially each household record will be weighted by the inverse of the selection probability. This weight will depend upon whether the data for the unit were ultimately collected by mail, telephone, or personal visit. These "base weights" would then be adjusted for nonresponse, as a means of accounting in the estimation system for those non vacant households that failed to provide data through any of the mail, telephone, or personal visit modes. For cases of missing item-level data from respondents, imputation would be used to compensate for the missing data. Finally, some sort of poststratification is envisaged, to ensure that the survey weights agree, at some level of geography, with accurate independent estimates of the population size. Since the exact nature and benefit of the poststratification procedures are unclear at present, in its current research the Census Bureau is evaluating the reliability of the ILF estimates under the conservative assumption that there will be no benefit from such a procedure.

4. Acceptance of Continuous Measurement Data Products

Continuous measurement would provide users with more current data (for all but the smallest geographic areas) and with greater frequency than data collected every ten years. A major uncertainty for assessing the costs and benefits of a continuous measurement program is the value of more timely data. More frequent estimates could be especially valuable to those making decisions about the distribution of funds or who are concerned with measuring characteristics of populations that change considerably during a ten year period.

Replacing single-point-in-time data for small areas with five-year moving averages may be troublesome for some users. Many census long form data users may be unfamiliar or uncomfortable with the concept of moving averages and may have to re-examine their use of the data. The Census Bureau has begun to work with census users at some federal agencies, in particular the Department of Transportation and the Department of Housing and Urban Development, to acquaint them with the type of data product that would be produced by a continuous measurement program. Discussions with data users in state and local governments and with private sector users are also planned.

It may be difficult to interest census data users without a data product, since continuous measurement is new and different. It is easy for users to see what is lost by dropping the long form, but the benefits of a new program or new methods may not always be obvious. As mentioned, the Census Bureau will begin a random digit dialing (RDD) telephone survey test of 2,000 households per month in four sites starting November 1994 and continuing for six to nine months. One purpose of the test is to provide demonstration files of cumulative estimates to distribute to census data users. The estimates will be accumulations of cross section or snapshot surveys. The RDD survey will be followed in October, 1995 by a Master Address File (MAF) based test survey in four sites of at least 4,000 households per month and will provide data users additional experience using cumulative estimates. The demonstration files are intended to get user reaction to moving averages and to determine the level of demand and acceptance of continuous measurement data. In addition to developing simulated data products, the test will also help to identify and define operational issues. While a simulated data product will play a major role in measuring potential demand and acceptance of a continuous measurement program, it is also important that the Census Bureau continue with a vigorous program of explaining and discussing cumulative estimates. Although a change to continuous measurement requires considerable methodological and operational changes, the major question is whether it meets the requirements of small-area data users. Will data users regard more timely cumulative estimates as superior or inferior to single point-in-time data once per decade?

Ultimately the question that needs to be addressed is the extent to which continuous measurement improves accuracy, relative to its cost. Since continuous measurement estimates will be based on the same sample size, for each five-year moving average, as the census long form sample data estimates were based in 1990, the sampling errors for these two different types of estimates will be comparable. If continuous measurement estimates are subject to so much less bias than once-per-decade estimates, on average across the decade, that their total errors are substantially smaller then it is likely that

continuous measurement will be cost effective. If the bias that results from outdatedness over time for once-per-decade estimates is modest compared to the standard error of sampling, then continuous measurement will offer little relative advantage. Clearly the situation will vary with the level of geography and the characteristics being estimated. At broader geographic levels, sampling errors will be small, and hence the biases that result from outdatedness will be relatively major. Certain characteristics of geographic areas change relatively slowly over a ten-year period, for most geographic areas (for example, housing characteristics), and for these kinds of estimates the benefits of continuous measurement will be modest in most cases. To reach a conclusion about the benefits of continuous measurement, it is important to weigh the importance of accuracy (i.e., mean square error) across different levels of geography, and across different kinds of estimates. It is also important to consider what value is to be placed on a system that would make substantial improvements in accuracy for a few areas (those growing rapidly, for example), but might not improve accuracy greatly in the great proportion of geographic areas.

5. Evaluation of Costs and Benefits

Cost estimates of operating a continuous measurement program and the potential savings from eliminating long form questions are not yet well defined. Various assumptions relating to such matters as the cost of frame maintenance, response rates, follow-up effort required, and the percentage of interviews completed using various data collection modes (mail, telephone and personal visit) are based on very limited knowledge. The prototype tests could be very helpful in refining assumptions used for cost estimates and determining the direct costs of various operations. Estimates of the potential cost savings from eliminating the long form from the decennial census also need to be refined. The consequent reduction in respondent burden is likely to yield a modest increase in mail response rates, thus reducing follow-up costs. Elimination of long-form items should also make it easier and less costly to collect data during nonresponse follow-up.

The Census Bureau is committed to building a continuously updated Master Address File (MAF) regardless of the final 2000 census design. This activity could conceivably grow into a cooperative effort involving other federal, state, and local agencies. A continuous measurement program would require stronger updating of the MAF. With a traditional decennial census, the computer processing of Postal Service delivery files would be done regularly, probably once every year or two. Some clerical resolution of the postal files for decennial census use is scheduled to begin in 1995 and will continue through the rest of the decade, but some of the updating activity could be postponed until immediately prior to the decennial census. With continuous measurement, MAF updating would be done quarterly. How should the costs of developing and updating the MAF be allocated to continuous measurement, to the census, and possibly to other statistical activities? This question needs to be resolved to determine fully the continuous measurement costs.

In addition to providing more timely data, there are additional potential direct benefits of a continuous measurement program:

1. Improved data quality from maintaining a continuous data collection operation which would permit developing and maintaining a well trained field and operations staff. With a continuing operation such as would be needed to conduct a continuous measurement program, repeated opportunities are available to refine and improve the design, field and data processing procedure, survey instruments, and estimation procedures. The continuing presence of an experienced staff and an established operation would lead to a greatly reduced risk (compared to a once a decade collection) that a serious unforeseen problem will arise, and introduce efficiencies into the operations that are not possible with a major effort mounted once a decade.

2. Ability to maintain a continuous presence in local areas over the decade, which enable those concerned with response and coverage improvement to conduct more effective outreach programs and improve public response to the census.

3. Provide a more conducive environment for use of administrative records by providing the opportunity for periodic checks on the quality of administrative records to be incorporated in the program.

Several additional features that a continuous measurement program might offer must be seriously considered in evaluating its benefits. First, there is the potential use of the Master Address File as a frame for other demographic surveys. This might apply to existing surveys, or perhaps even more likely, to potential future surveys for which the cost of developing a frame would make them unaffordable. A second feature would be the use of the ILF survey itself as a screening device for rare populations, which would be subsequently surveyed at a later time via telephone or personal visit. Again, this feature would make feasible surveys that otherwise would have prohibitive screening costs. A third way in which a continuous measurement program might add a capacity to collect data efficiently is through the inclusion of supplements as part of the survey instrument. A supplement could be included for one month only, to give estimates for a particular topic of interest at a broad geographic level. The same supplement could be repeated

annually or at longer periods to give national and major subnational estimates of change over time. Alternatively the same supplement could be repeated for several successive months to give estimates at finer geographic levels.

In using the continuous measurement program to collect supplementary data, or as a screening device, care would be needed to ensure that the presence of these extra components did not negatively affect the response rates, or change the respondents answers to the core questions, thus affecting movement-over-time estimates for these core components.

The Master Address File itself might constitute a high quality cost-effective frame for other household surveys. The Census Bureau conducts a number of periodic household surveys--for example, the Current Population Survey, the National Health Interview Survey, and the Survey of Income and Program Participation--as well as one-time surveys, and the benefits of having up-to-date Master Address File as a sampling frame for these surveys are potentially great. Depending upon the legal requirements with regard to the confidentiality of the Master Address File data, perhaps this frame could be used by other federal statistical agencies for their own surveys (not conducted by the Census Bureau). If Title 13 of the United States Code is amended to permit access to allow sharing of address lists with federal, state, and local officials, as has been proposed (see the March 1994 issue of Census 2000 Update), then not only would the use of the Master Address File as a frame be possible at a federal level, but it could also be used by state and local agencies for conducting surveys.

Other benefits and cost savings exist that will be difficult to quantify, even should a continuous measurement program become fully operational. These include improved data quality resulting from a greater ability to develop and retain a well-trained field and operations staff, a uniform work load for managing and controlling operations, and more cost-effective use of hardware and software for data collection and management systems. Thus, a continuous measurement program has the very real potential to enhance the data collection capability of the Census Bureau (and the federal statistical system more generally) to include areas that heretofore have not been practical because of cost considerations. Although it is difficult to reflect this as a cost saving resulting from continuous measurement, nevertheless this important potential benefit must be considered when evaluating the program.

6. Will Continuous Measurement Improve the Decennial Count?

As discussed in Section 1, it has been argued that the decennial census could obtain a more complete count of the population if the questionnaire were limited to only those questions needed to satisfy the legal minimum requirements (about which there does not yet seem to be complete agreement). A continuous measurement program, by serving as an alternate source for the long form information, would thus contribute to a more accurate (and less costly) decennial census.

To what extent does available evidence support this argument? Unfortunately, the answer is not clear. While the cost savings associated with dropping the long form have been debated extensively, as have the costs of conducting a continuous measurement program, the impact on the decennial count appears to have received less study.

Results of tests conducted as part of the 2000 research program and elsewhere show clearly that, in general, the shorter the form the higher the mail-back response rate. By itself, however, this result does not necessarily lead to a more complete count. There is, for example, no evidence that within-household undercoverage (about half of total underenumeration) is reduced, nor is it clear that the traditionally hard-to-enumerate populations will be substantially better counted.

Furthermore, even a noticeable difference in response rate for the long from compared to the short form may not have a substantial impact on the overall mail response rate, since only a minority of households receive a long form. For example, if the mail return rate were to be 5 percentage points lower than for the short form, and if only 16 percent of households receive a long form (as was the case in the 1990 census), then the mail return rate would only be 5 x 0.16 = 0.8 of a percentage point lower than the mail return rate that would be achieved if only short forms were to be collected.

A comparative analysis of mail return rates from the 1990 census suggests that dropping the decennial long form would yield a trivial improvement in census coverage (Keeley, 1993), and there is little evidence that differentials in coverage would be affected. Therefore, the presence of the long form does not appear to diminish the accuracy of the decennial population totals used for reapportionment, redistricting, and resource allocation.

It is plausible and reasonable to assume that coverage improvements could be achieved by diverting some of the resources saved from the elimination of the long form to improved coverage and follow-up activities. The extent of such improvement, however, is difficult to estimate, and there does not appear to have been any definitive research in this area.

In summary, then, while it is possible that the decennial count could be improved by dropping the long form, it seems that the extent of such improvement is likely to be small, with little impact on differential undercoverage.

7. Impact of Non-Census Environment on Response Rates and Data Quality

One of the most important questions which must be answered in assessing the feasibility of a continuous measurement program as an alternative to a census long form is the extent to which response rates (ultimately, costs and data quality) will be affected by the fact that the program would be conducted in an environment vastly different from that of the census.

The decennial census is carried out with extensive publicity, some generated through the public communications activities of the census program itself (advertising, outreach efforts, etc.), and even more by the fact that the census is a highly newsworthy event which attracts a great deal of media attention. Such extensive publicity generates a high level of public awareness, which serves not only to stimulate initial mail-back response rates, but equally important, to motivate census field staff and make it much easier for them to conduct follow-up activities. Both respondents and field staff are aware that they are participating in the decennial "national portrait", and even if a general decline may be occurring in the sense of civic responsibility that motivates participation, there is no doubt that the census benefits greatly, in terms of response rates, costs and data quality, from the publicity surrounding the program.

A continuous measurement program, on the other hand, would be conducted in the virtual absence of such publicity. While the launching of such a major new survey could be expected to generate some significant attention, it is highly unlikely that media attention would be sustained, and it would be prohibitively expensive to conduct continuing national and local advertising campaigns in support of the program.

There is little useful evidence now available on which to base estimates of the impact of the "non-census" environment on response rates, costs, or data quality. The tests conducted to date in conjunction with the 2000 Census have little direct relevance to this issue, nor do other government or commercial surveys yield much insight. A recent (November, 1993) national (long form) census test in Canada, where recent mail-back response rates have been considerably higher than in the U.S., produced an initial mail response rate of less than 50 per cent. This test, however, differed in many respects from the proposed program of continuous measurement.

One factor that might have a substantial impact on response rates is whether response to the ILF survey is mandatory, as for the decennial census, or voluntary, as for the Census Bureau's household surveys. It may be useful to establish whether response to the ILF survey could be ruled mandatory under current or revised statutes (Title 13, U. S. Code). A well-implemented, mandatory survey might achieve mail return rates similar to those obtained for redesigned long forms in recent Census Bureau experiments.

What is clear at this time is that it is not possible to assess the credibility of the response rate assumptions used in the current Census Bureau plans. As a result, cost estimates are subject to a very wide range of uncertainty, as are estimates of data quality, since there is no information available on which to assess inevitable differences in response patterns among various important sub-populations. Given these uncertainties, which are recognized by Census Bureau staff, the conduct of a research and testing program that will enable the development of sufficient evidence on which to make informed assessments on issues related to costs and data quality is essential in the development and evaluation of a continuous measurement program.

8. Effects of Demands to Make Changes in Form and Content Over Time

A concern often raised during discussions of continuous measurement as an alternative to the census long form is that pressures for changes in the continuous measurement content and design throughout the decade (whether through budget reductions or emerging new data needs) would lead to a loss in comparability over time. While such pressures will be inevitable, it does not necessarily follow that they will be unmanageable or that the possibility of such problems should weigh heavily in the decision on whether or not to proceed with the continuous measurement research and testing program.

Significant fluctuations in budget allocations to a continuous measurement program would create serious difficulties for the planned output of such a program, but should such fluctuations occur, changes in sample size (rather than content) would be the most likely response. Such changes would have an impact on geographic and sub-population detail and accuracy, and would reduce the frequency of the output.

Demands for changes in content would most likely take the form of requests for new questions or additional detail on existing topics. The response to such demands could most easily be handled through the conduct of supplementary surveys, discussed in section 5.

9. Conclusion

The above discussion should serve to indicate the potential that surrounds an evaluation of the merits and feasibility of a continuous measurement program to collect small area sample data. Many aspects of this program need to be developed and evaluated in tandem, and a great deal more needs to be tested and learned on all fronts before rational decisions about the prospects for such a scheme can be reached. In addition to the question of whether a continuous measurement program such as the one proposed has merit over the long term, there are additional questions relating to the viability of introducing a continuous measurement program of sufficient quality and scope in short enough time to act as a viable replacement for any small area data collection (i.e. a long form) as part of the 2000 census.

Finally, a continuous measurement program would not operate in isolation. It would be an integral part of the Federal Statistical System. As such, its relationship to, and impact on, other federal data collection efforts, (in particular, the various ongoing household surveys) must be considered.

REFERENCES

Alexander, C.H. (1993). *A Continuous Measurement Alternative for the U.S. Census.* Internal Census Bureau document CM-10.

Alexander, C.H. (1994) *Progress on the Continuous Measurement Prototype.* Internal Census Bureau document CM-12.

Committee on National Statistics (1993), *Planning the Decennial Census: Interim Report.* Panel on Census Requirements in the Year 2000 and Beyond, Committee on National Statistics, National Research Council, National Academy Press: Washington, D.C.

Committee on National Statistics (1994). *Counting People in the Information Age.* Panel to Evaluate Alternative Census Methods, Committee on National Statistics, National Research Council. Washington, DC: National Academy Press.

Goldfield, E.D. (1992). *Innovations in the Decennial Census of Population and Housing: 1940-1990.* Paper prepared for the Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, National Research Council.

Herriott, R.A., Bateman, D.V., and McCarthy, W.F. (1989). The decade census program -- a new approach for meeting the nation's needs for sub-national data. *Proceedings of the Social Statistics Section.* Alexandria, VA: American Statistical Association.

Horvitz, D.G. (1986). Statement to the Subcommittee on Census and Population, U.S. House of Representatives (May 1). Research Triangle Institute, Research Triangle Park, N.C.

Keeley, C. (1993). *Could the Census Bureau Reduce the Undercount by Not Using a Long Form?* Bureau of the Census, U.S. Department of Commerce.

Kish, L. (1981). *Using Cumulated Rolling Samples.* No. 80-52810. U.S. Government Printing Office, Washington, D.C.

Kish, L. (1990). Rolling Samples and Censuses. *Survey Methodology.* 16: 63-71.

Lee, M.M. (1993), *Legal Issues for Census 2000.* Congressional Research Services Report 93-177-A. Washington, D.C.: U.S. Government Pointing Office.

Sawyer, T. (1993). *Rethinking the Census: Reconciling the Demands for Accuracy and Precision in the 21st Century.* Transcript of Congressman Sawyer's statement at the Census Bureau Research Conference on Undercounted Ethnic Populations, May 5-7, 1993, Richmond, Virginia.

U.S. Bureau of the Census (1994). *Census Content and Questionnaire Design:* Bureau of the Census, U.S. Department of Commerce.

U.S. House of Representatives (1993). Report of the House Committee on Appropriations (June 24, 1993). H.R. 2519 Department of Commerce, Justice, and State, the Judiciary, and Related Agencies Applications Act, 1994 (Pub. L. 103-21).

EXHIBIT 1

Attachment A

ACCELERATED MAF-BASED CONTINUOUS MEASUREMENT
Data Collection Activities

FY	Data Collection Activity	Objectives
1994	Research, planning, outreach only.	o Win over a few key Federal users o Contact non-Federal users o Remove feasibility doubts o Get commitment to 1996 testing.
1995	RDD Test with 2000/month total in 3-4 sites, starting November 1994. Convert to split-sample questionnaire test in July 1995. Small mail pretest.	o Get demonstration file of cumulative estimates o Test alternative versions of questionnaire o Get user acceptance of testing/decision process o Get commitment to FY 97 funding and decision process
1996	MAF-based test with at least 4000/month total in 4 sites, starting October 1995.	o Better demonstration file of cumulative estimates o Develop/test field procedures o Get user input and decide whether to proceed further o Decision regarding 2000 long form, assuming FY 97 results successful o Get commitment to FY 1998-2002 funding conditional on decision process
1997	MAF-based "development survey" for Congressional-District-level estimates, full speed in January 1997. Rural sample clustered in PSUs.	o Demonstrate actual procedures o Produce actual high-level estimates o Measure coverage, quality of estimates o Close scrutiny of 1995, 1996, 1997 data by all users. o Final decision regarding 2000 long form in December 1997
1998	Expand MAF-based sample size; change procedures and questionnaire to fix problems found in FY 1997. Better rural spread.	o Final content determination o Final procedures o Further evaluation of quality o More actual high-level estimates
1999	Full MAF-based system. Complete rural spread.	o Collect small-area data to replace 2000 long form