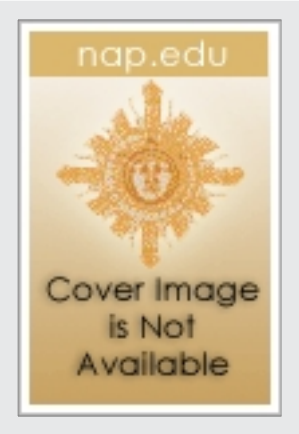

Experimentation and Testing Plans for the 2010 Census: Letter Report

DETAILS

13 pages | 8.5 x 11 | PAPERBACK
ISBN 978-0-309-13651-8 | DOI 10.17226/12607

AUTHORS

Panel on the Design of the 2010 Census Program of Evalutions and Experiments; National Research Council

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

500 Fifth Street NW
Washington, DC 20001
(202) 334-3096
Fax: (202) 334-3751

Panel on the Design of the 2010 Census Program of Experiments and Evaluations
Committee on National Statistics
Division of Behavioral and Social Sciences and Education

February 19, 2009

Mr. Thomas L. Mesenbourg
Acting Director and Deputy Director
U.S. Census Bureau
4600 Silver Hill Road
Washington, DC 20233

Dear Director Mesenbourg:

This letter relates to plans for tests and experiments planned for the 2010 census. We write to call your attention to several time-sensitive concerns: (1) three crucial topics that should be included in the experimentation during the 2010 census, (2) testing plans preliminary to the census; (3) the retention of 2010 census data, and (4) the designs of the experiments currently planned for 2010.

Background

The Panel on the Design of the 2010 Census Program of Evaluations and Experiments (CPEX) has a broad charge:

> . . . [to] consider priorities for evaluation and experimentation in the 2010 census. [The panel] will also consider the design and documentation of the Master Address File and operational databases to facilitate research and evaluation, the design of experiments to embed in the 2010 census, the design of evaluations of the 2010 census processes, and what can be learned from the pre-2010 testing that was conducted in 2003–2006 to enhance the testing to be conducted in 2012–2016 to support census planning for 2020. Topic areas for research, evaluation, and testing that would come within the panel's scope include questionnaire design, address updating, nonresponse follow-up, coverage follow-up, unduplication of housing units and residents, editing and imputation procedures, and other census operations. Evaluations of data quality would also be within scope. . .

Pursuant to this charge, the panel transmitted an interim report providing general priorities for the CPEX program to the Census Bureau in late 2007 (National Research Council, 2008) and plans to issue a final report in fall 2009.

The panel met most recently on November 10–11, 2008. At that meeting, Census Bureau staff briefed the panel about the topics that it had chosen for inclusion in the 2010 CPEX program and presented the outlines of the designs for the experiments to be included in the 2010

census. On the basis of those briefings and subsequent discussion, and given the relatively late timing of our final report in the census experimentation planning cycle, the purpose of this letter is to continue to fulfill our charge by providing timely analysis and recommendations for the CPEX program.

Experimentation During the 2010 Census: Missing Topics

A key objective of our interim report (National Research Council, 2008) was to suggest priority topics for experimentation during the census. In particular, we urged that the topics chosen for experimentation have a direct bearing on visions for the 2020 census (however preliminary) so that they can serve as a first step for research in the intercensal period. We also explicitly recommended that the 2010 experiments be chosen to examine issues with the potential to achieve substantial cost reductions or important improvements in data quality in 2020.

In November 2008, the panel was informed that the Census Bureau has chosen topics for four experiments to be conducted during the 2010 decennial census: (1) a nonresponse follow-up contact strategy experiment, (2) a privacy notification experiment, (3) an alternative questionnaire experiment, and (4) a deadline messaging and compressed schedule experiment. We are deeply concerned that although the topics selected by the Bureau are of interest, they are not grounded in a vision for 2020, nor are they directly linked to cost or data quality concerns. At the same time, we are concerned that two topics with strong potential effects on cost and quality and overall importance for 2020 that we discussed in our interim report are absent from the Bureau's experimentation plans: Internet data collection and the use of administrative records. We reemphasize that these two areas of research are critically important. In addition, we believe that a very different alternative questionnaire experiment—one that tries multiple approaches to improve collection of census residence information—would be invaluable for the future of census questionnaire design.

Internet Experimentation The use of the Internet for data collection in the decennial census presents important opportunities for cost reductions and improvements in data quality. These include cost savings through the reduction in the number of forms that have to be scanned or keyed for data entry, reduction in the processing of requests for mailing of foreign language questionnaires, and savings in field work as a result of more prompt receipt of individual data. Use of the Internet may also yield quality improvements through easier access to foreign language questionnaires and online editing of census responses. Importantly, the use of online response would avoid the social cost of the Census Bureau's appearing to be out of step with modern data collection and computing environments.

An experiment in the 2010 census would provide a unique opportunity for examining the use of the Internet for decennial census data collection. A key issue that needs to be explored in an experiment is how large a fraction of the population can be induced in a census environment to use the Internet as a response option, while not at the same time greatly increasing the possibility of disclosure or incurring other security problems. Therefore, we strongly recommend a 2010 census Internet response experiment to help determine ways to increase the likelihood of Internet response in 2020 and possibly also learn how to minimize any associated negative effects. This test should include a "push Internet" option as one of the experimental treatments whereby the initial mail contact strongly encourages Internet response, perhaps even by

excluding a paper questionnaire from that initial mailing. Such an experiment could also address the quality of the data collected through the Internet, including for those requiring foreign language questionnaires for whom the Internet may provide a convenient multi-language option.

We recognize that the basic steps to implement an Internet experiment in 2010 are nontrivial: the design and testing of an online version of the census questionnaire, the development of protocols that protect census respondents from disclosure of information, and the integration of online returns with other census operations. However, the panel is confident that the challenges can be overcome, even within a tight time frame, as they were when the Census Bureau added a limited online response option in 2000. In addition to the Census Bureau's own experience with Internet questionnaire development in the 2000 census, the experience of other countries in developing security protocols for online census response (including the 2006 Canadian census) can be tapped as the Census Bureau develops privacy safeguards for online response in planning such an experiment.

Use of Administrative Records Administrative records offer substantial potential for both census cost reduction and quality improvements. Administrative records could be used to dramatically reduce the cost of nonresponse follow-up and improve the quality of the resulting data collected by avoiding inaccuracies in "last resort" enumerations (often supplied by proxy respondents, such as neighbors or landlords) and by providing higher quality information than is currently supplied by whole-person and whole-household imputation. (An admittedly radical eventual possibility for the use of administrative records would be avoidance of nonresponse follow-up altogether for a large percentage of U.S. households.)

In addition, administrative records could be used to target the implementation of census processes. A key example is that administrative records could identify areas in which the Master Address File (MAF) is deficient, by basing that determination on the difference between the address counts from a merged list of addresses from administrative records and the counts from the MAF, and therefore in need of an address canvass check prior to the decennial census. This approach could dramatically reduce the costs of the currently 100 percent application of the address canvassing operation. One could also use the discrepancy between a household count from the census and that from administrative records to prioritize the implementation of coverage follow-up interviews. Finally, administrative records could be used to assist in reducing the field work in following up nonmatching cases of the P-sample in coverage measurement.

Although wide-scale use of administrative records to substitute for nonresponse follow-up would almost certainly require a change in legislation, the potential benefits of increased use of records in census processes should be studied in order to estimate the extent to which such changes would be economically and statistically desirable. Given that the use of administrative records in such a manner provides one of the few opportunities to substantially reduce census field costs in 2020, it deserves serious attention in the planned 2010 experiments.

It is important to note that most of the above possibilities for research on administrative records might be properly considered priorities for "evaluation" rather than "experimentation" since they would not require additional or special field data collection. (They would, however, require the careful retention of household-level census process data, such as we recommend below.) Yet although a great deal about the utility of administrative records can be learned from post hoc study of data retained during the census, there are potentially useful possibilities for limited, experimental field work in 2010. For instance, with regard to the use of administrative

records as a substitute for late-stage field enumeration, one possible experiment would involve variations in nonresponse follow-up or coverage follow-up protocols under which the number or format of follow-up interviews depended on administrative records information (either on an individual household basis or on an area basis). Such an experiment would involve a significant expansion of the nonresponse follow-up contact strategy experiment (discussed below).

Though "administrative records" in the census context are generally thought to be national-level constructs—drawing information from, for example, Social Security Administration registers—a complete evaluation of records-based methods should also assess the quality of the records maintained by "group quarters" facilities, such as prisons, health care facilities, and college residence halls. Because these facility records were used by census enumerators to count about half of the group quarters' population in the 2000 census, the National Research Council (2006:Table 7-1, pp. 238-240) suggested that the Census Bureau "undertake a continuing research effort to assess the accessibility of facility records at group quarters facilities and to determine whether the existing data systems meet census data collection needs." We endorse this suggestion as it is an essential step to assessing the possibilities for using administrative records to supplement or, as necessary, replace traditional enumeration in group quarters. Assessing the alternative or "home" address information available from facility records is also critical to addressing such long-standing questions as the degree to which college students are counted at both their schools and their parental homes and whether it is feasible to define a "home address" for persons under correctional supervision.

Census Residence The 2010 census provides a uniquely valuable setting for a comprehensive experiment involving alternative approaches to the current residence rules. The Census Bureau's proposed alternative questionnaire experiment for 2010 does include one treatment group for gathering a limited amount of information on residence (see below). However, given that unclear residence rules and interpretations were likely a major source of census coverage error (both omission and duplication) in the 2000 census (National Research Council, 2004), the Panel on Residence Rules in the Decennial Census (National Research Council, 2006) suggested various alternative approaches to collecting information on census residence. In particular, that panel's report proposed a major change from the Census Bureau's traditional approach of relying on a dense set of instructions at the start of the census form to one of asking a set of guided questions that breaks the large cognitive task of deciding one's household composition into smaller pieces. At that panel's urging, the Census Bureau tested a preliminary version of a "worksheet" approach to the residence question in 2005, yet no further work on residence is planned in 2010.

The single treatment group in the proposed alternative questionnaire experiment—anchored to one of the coverage probe questions—falls short of the general "any residence elsewhere" query that the National Research Council (2006) recommended be asked of the general population in a 2010 census experiment and asked of all group quarters (e.g., medical facilities and college housing) residents in the 2010 census itself. The current plans for this limited experiment also do not appear to include the follow-up activities needed to make best use of whatever information might be gained. The proposed single treatment group also falls short of the 2006 report's suggestion to experiment with a *de facto* or "current residence" question—and add a corresponding *de jure* or "usual residence" question to the American Community Survey—so that differences in estimates between the two programs due to their differing residence standards could be assessed. Innovative (and more accurate) handling of residence

concepts is clearly a research question for which several alternatives need to be tested, and subsequently refined and retested, in order to achieve substantial gains over the Bureau's current approaches.

These three research areas—Internet data collection, the use of administrative records, and questionnaire redesign for residence rules—are ones for which important benefits could be obtained through increases in census data quality or decreases in census costs or both. In the panel's assessment, the 2010 CPEX program should include work on these topics in order to ensure early progress in the 2020 census testing cycle. Therefore, we strongly urge that these topics be included as subjects for experiments in conjunction with the 2010 census.

Systems Testing and Simulation Prior to the 2010 Census

The panel is concerned that the Census Bureau's operational test plans for the 2010 census are insufficient. We are particularly concerned with the Bureau's capacity to identify potential failure modes in the field data collection components of the 2010 census process. We appreciate that the Census Bureau has had to substantially revise its plans for decennial census nonresponse follow-up. Initial plans to use handheld computers for nonresponse follow-up and to have the operational control system for field data collection developed by a contractor have been dropped in favor of a return to a paper-based nonresponse follow-up operation and a return to an operational control system for field data collection that will be developed in house (presumably by revising the system developed for the 2000 census).

Given the complexity of conducting the decennial census, it has long been deemed essential to have a complete test "dress rehearsal" two years prior to the census so that flaws can be detected and corrected. Given the need to redesign the field data collection plan at this late stage, the census dress rehearsal conducted in 2008 was essentially limited to a test of the mailout/mailback portion of the census process, with no testing of the nonresponse follow-up, coverage follow-up operations, or many other component processes.

The Census Bureau acknowledges that the dress rehearsal provided an inadequate test of the 2010 census processes. As a remedy, it has scheduled a number of small field tests of various components and sub-systems of the census process chain to attempt to identify as many potential flaws as possible prior to implementation. However, given that the operational control system for the field data collection system will not be ready until the summer or fall of 2009, the Census Bureau has decided against a comprehensive test of the entire field data collection process due to the lack of time to design and carry out such a test.

The panel believes that this testing strategy puts the Census Bureau in an extremely risky position should there be flaws in the census process that involve interactions of the many components and subsystems. Testing the interfaces between individual components of a system (e.g., $A \rightarrow B$, $B \rightarrow C$, $C \rightarrow D$) can produce useful information and detect unseen problems. But the Bureau's testing plan creates risks by not adequately testing subsystems (e.g., $A \rightarrow B \rightarrow C$) or complete systems. Errors at this level may not be evident in any single component test but could result in major delays and impair data quality.

Concern over the lack of time or resources to conduct a more comprehensive test is understandable, but it does not override the compelling argument for carrying out such a test. The Census Bureau needs to perform as full and realistic an operational test of all nonresponse follow-up systems as possible. The consequences of failure to identify substantial problems in the interfaces between system components could be dire, ranging from moderate to severe

impacts on the quality, costs, and timeliness of census counts for important purposes like redistricting and allocation of funds.

The panel strongly recommends that the Census Bureau try to fit into its schedule a comprehensive test of the entire operational control system for field data collection as soon as feasible after plans for this system become available. We recognize the enormous constraints in planning and accomplishing such testing. Because of these constraints, it may well be necessary in the overall testing to simulate portions of the process based on the specifications for information flows at the interface between component parts of the process. If such simulation is judged to be necessary, then additional field testing of the simulated components of nonresponse follow-up should be carried out.

Ideally, tests should be conducted in enough time to detect—and correct—any problems. But if time is too short to allow for a full cycle of test and correction, earlier detection of defects or inefficiencies can still be vital. Even if a flaw is discovered too late to be addressed in a pre-tested, systematic way, some contingency planning will likely be able to greatly reduce any negative consequences for the census itself.

Retention of Data

Since 1985 several National Research Council panels on the decennial census have called for the development of a "master trace sample" database. Such a database would retain the crucial elements of the census procedural history for a sample of addresses to support census evaluation studies. A version of a master trace sample was constructed by the Census Bureau following the 2000 census (Hill and Machowski, 2003). This database supported a small number of studies (e.g., Bentley and Tancredo, 2005; Tancredo and Bentley, 2005; West et al., 2005) that began to realize some of the substantial research potential that such a database could provide.

Our panel's interim report recommended that "the Census Bureau should initiate efforts now for planning the general design of a master trace sample database and should plan for retention of the necessary information to support its creation" (National Research Council, 2008:Rec. 5). To address the efficacy of less common procedures on small subpopulations, a large sample is clearly needed; we also note that given the greatly decreased cost of computer storage and memory, it may now be possible to save and efficiently access the entire procedural history for the entire country. Whatever the sampling rate, it is critical to retain sufficient data, preserving all relevant linkages, so that the result supports the examination of how the decennial census processes functioned for various subpopulations and domains.

As an example, it is important to retain the information as to which addresses on the MAF were added or deleted by which census address improvement operations. Furthermore, given that many fields of the various system files are overwritten continuously during the census, this means that these data archives should retain snapshots of files that will change during the course of census operations, and this should be provided for as frequently as needed. This data archival effort needs to include all parts of the census process, including address list development, nonresponse follow-up, coverage follow-up, group quarters enumeration, data capture and data treatment, and coverage measurement. In addition, it is vital that the schema used in retaining these data be carefully documented so that it is known precisely what is saved in each data field.

Given the rushed development of the operational control system for field data collection, we are especially concerned that provisions be made for retaining data relating to that part of the

census. We do not believe that providing for this additional functionality in the operational control system for the field data collection will add appreciably to the current challenge of developing such a system in time for the 2010 census. Furthermore, by guaranteeing access to this information, the Census Bureau would ensure that it could carry out evaluations that would guide the Bureau towards a more effective and cost-efficient design for the 2020 census. Therefore, we recommend that—as systems for the 2010 census are finalized by the Census Bureau and its contractors—appropriate archival outlets be created for all systems, including components of the field data operational control system, so that the relevant data to construct a master trace database or "audit trail" of census processes are retained. Experts in automated audit processes could provide assistance to the Census Bureau in implementing a master trace system.

Designs for Currently Planned Experiments

Although we recommend the addition of three topics for experimentation, the Census Bureau's chosen topics for 2010 experiments do concern issues that may be worth pursuing in addition to our recommended ones. However, three of the four census experiments, as currently outlined, suffer from important defects that will limit their effectiveness. Moreover, the Bureau has not carried out explicit studies of the statistical power of these experiments given their proposed designs. We recognize that the clustering inherent in some of the experimental designs complicates the development of such estimates, but it is also the reason that careful estimates of power are necessary. For each experiment, the Census Bureau needs to undertake a study of the statistical power of the design against reasonable alternatives based on anticipated effect sizes. This should be done not only for national-level comparisons, but also for any relevant subgroup comparisons.

Some of the experiments also do not seem to give appropriate attention to "targeting" or oversampling respondents from relevant sociodemographic groups (or geographic areas with large concentrations of such respondents). Not only does lack of targeting reduce the power of those experiments, but it also hinders the ability to learn more about the response by stratifying the analysis by subgroup.

The Nonresponse Follow-up Contact Strategy Experiment The question of interest in this experiment is the impact on census costs and data quality of reducing the number of attempts made in nonresponse follow-up from a maximum of six to either four or five. As currently planned, the experiment will be carried out in three local census offices, comprising about 40,000 housing units. For each office, two treatments and the control will be randomly allocated to crew leader districts, where all enumerators in a district will use the same questionnaire (which provides space for a maximal number of enumeration attempts) but will receive different instructions about how many callbacks to make. To assess the treatments and control, comparisons will be made of the resulting impact on census data quality, measured by the rate of proxy response, the distribution of response outcomes, the item nonresponse rate, and measures of form completeness. The Census Bureau staff have expressed a concern as to whether the findings would be generalizable from the three local census offices, and asked the panel for assistance in selecting local census offices for this experiment. However, our current overriding concern is whether data from only three local office areas can ever be sufficiently generalizable.

In addition to questions about generalizability and statistical power, the panel questions whether the likely reduction in field data collection costs will be sufficient to justify the

allocation of resources for an experiment during the 2010 census. The likely impact on census costs might be fairly modest. In the November meeting, the panel suggested that the cost reduction could be estimated on the basis of the frequency of enumerations in 2000 that were successful on the fourth or fifth attempts. The Census Bureau argued that such estimates are misleading due to infrastructure changes that occur during the taking of the census, such as the laying off of enumerators, consolidation of work, and other changes. The panel countered that estimates based on an analysis of 2000 census data, while somewhat flawed due to such changes, would still provide a sense of whether the potential reductions in field costs would be large enough to justify a separate experiment during the 2010 census. Based on such estimates, if the cost reduction seems likely to be, at best, modest, the experiment should be eliminated or redesigned to include assessment of even fewer enumeration attempts or the use of administrative records in lieu of field data collection.

In considering statistical power, 2000 data could have been used to estimate the percentage of housing units that first failed to return their mailed questionnaire, and then were enumerated in the 2000 census on either the fourth or fifth attempt during nonresponse follow-up. In doing so, it may be discovered that the effective sample size for this experiment is too small to provide sufficient power to identify important differences in the above data quality measures (unless such differences are strikingly large). If it is clear that the experiment will not have substantial power to detect reasonable changes to the census data quality measures, and if a two or three-fold increase in the number of local census offices would provide sufficient power, the sample size should be expanded. If no conceivable sample size can provide reasonable statistical power, the experiment would not be useful and should not be done.

One additional argument in favor of an experiment on this topic, if slightly broadened, is that there is an a distinct disadvantage of waiting until six responses are attempted. This disadvantage is that the lag between Census Day and the day of enumeration increases the number of movers and in general reduces data quality and increases the rate of erroneous enumeration. Assessment of this disadvantage, possibly in conjunction with the coverage measurement program, might be very useful.

The Privacy Notification Experiment The privacy notification experiment will assess the effect of a message on the cover letter of the mailing package containing the census questionnaire regarding the uses of census data and the possible use of administrative records. The experiment includes two panels of 10,000 sampled households each (plus a control group without such notification), chosen using strata based on levels of mail response in the 2000 census or in the American Community Survey. The assessment of the three wordings will use response rates, data quality measures, and monitoring of public reaction. The hope is to be able to have reasonable power to identify a difference in overall mailback rate of 1.8 percent. (A one percent reduction in mail response is estimated to cost the Census Bureau $90 million in 2010.)

The panel has three principal concerns with the current design of this experiment. The treatment panels vary only in the wording of one part of the notification message—"Your answers will be used for statistical purposes, and no other purpose" compared with "Your answers will only be used to produce statistics"—raising concerns about how informative the test will actually be regarding individual perceptions of privacy. Second, a longer, second section of the message is identical between the two treatment groups and hints at the possible use of administrative records:

> To improve census results, other government agencies may give us information about your household. The additional information we receive is legally protected under Title 13, like your census answers.

If the objective of the experiment is to assess privacy concerns, it would be beneficial to explore other wordings of this second part of the notification. Instead of a single test of a very limited set of alternative statements in 2010, it would be more useful for the Census Bureau to conduct a series of intercensal tests between 2010 and 2020 that would develop a broad sense of people's sensitivity to privacy concerns and use of administrative records. Such a research program should examine this for sociodemographic subsets of the population.

Another deficiency is that the Census Bureau is not using this opportunity to evaluate the implied tradeoff of the costs incurred from the freedom to use administrative records as a result of the inclusion of such a notification and the benefits from being allowed to do so. That is, while the privacy notification may have the effect of reducing mail response rates, it will at the same time allow for the use of administrative records to reduce costs and improve data quality, for example, by substituting for last-resort and proxy enumeration. Therefore, it seems reasonable to use this opportunity to determine the degree to which administrative records can reduce census costs and improve census data quality and whether such benefits offset the reduction in mail response and the associated increase in the costs of nonresponse follow-up. Possibly, this could be done through the separate administrative records experiment noted above, but bundling this as a single experiment may have some advantages, although it would increase the complexity of the currently planned experiment**.**

The Alternative Questionnaire Experiment There are three parts to the proposed 2010 questionnaire experiment: (a) a comparison of the complete set of questionnaire changes between 2000 and 2010, (b) an attempt to collect an alternative residence address based on answers to a coverage probe question, and (c) alternative formats for the collection of information on race and ethnicity. In part (a), 10,000 housing units will receive a 2000-style census questionnaire. Comparisons will be made to the distribution of responses to the full 2010 census to ascertain what changes between 2000 and 2010 are due to changes in questionnaire format. In part (b), 30,000 housing units will be administrated an alternative questionnaire that will permit respondents to specify a street address if they indicate that the person in question sometimes lives or stays at another location. In part (c), 30,000 housing units in each of 11 panels will be administered various questionnaire formats for the questions on race and ethnicity. Some of these will present slightly different versions of a combined race and Hispanic origin question (the 2010 census questionnaire itself presents them as separate numbered items). Other treatment groups respond to census advisory committee suggestions by permitting multiple and write-in answers to the Hispanic origin question or varying specific examples that are explicitly mentioned in the question (e.g., Taiwanese or Marshallese). It is planned that cognitive testing will be carried out in advance of the experiment to better refine the various alternatives. The forms will be mailed to a random sample of housing units, and initial nonrespondents will receive a replacement questionnaire that mimics the initial questionnaire. The goal of the experiment is not to identify specific alternative formats, but rather to learn more about the general formats that are preferred in order to fold this information into a longer term research program on questionnaire design.

The goals of parts (a) and (b) are not clear to the panel. Consequently, it is hard to judge whether the experimental designs and sample sizes are suitable and whether the experiments are likely to yield useful results. Although the sample size for part (a) may be sufficient to detect any economically important change in overall response rates between these two forms of the questionnaire, it may not be adequate if one wishes to understand how these changes are related to subgroups of the population, size of family, etc. The sample size is also not likely to be adequate if one is attempting to relate specific changes in response patterns to specific living situations, membership in demographic subgroups, etc. Otherwise, interpretation of any changes in response patterns will be limited due to confounding as a result of the several simultaneous changes to the questionnaire. As a result, the benefits for questionnaire design for 2020 will be reduced.

With respect to part (b) we are concerned about adequate power because it was unclear that 30,000 households would provide a large enough number of alternative addresses to be able to determine whether the inclusion of such a question on the census questionnaire would be able to substantially affect the need for the coverage follow-up interview or the accuracy of such an interview if it appeared to be needed. Therefore, some form of targeting—say of areas with a high frequency of seasonal second homes, or of people living in types of group quarters that frequently involve duplication—would be desirable. Second, it was not clear that this part included sufficient provision for gathering follow-up information so as to determine the usefulness of the additional question. That is, although the addition of any question on the census form has an associated cost of processing and a possible decrease in overall data quality, the inclusion of this question could produce higher quality responses as to census residence and/or it could also affect the frequency of coverage follow-up interviews or their accuracy. Therefore, it is important to include plans in the experimental protocol that would attempt to evaluate this tradeoff, since this should be key to making any decisions about the inclusion of such a question in the 2020 census questionnaire.

The race/ethnicity arms of this experiment (part c) involve fine distinctions in question wording that are most applicable to specific demographic subgroups. In particular, a major emphasis in this section is on Hispanic respondents. Therefore, this experiment would greatly benefit from any efforts to target the delivery of the questionnaire to areas with a larger percentage of Hispanic residents. In addition, given the increased use of bilingual questionnaires in the 2010 census to facilitate response for essentially the same population, it would be useful to extend this experiment to examine the impact of such changes on a bilingual version of the census questionnaire.

The Deadline Messaging and Compressed Schedule Experiment The key question of this experiment is whether the rate of mail response could be increased as a result of the use of deadline messaging (namely, the use of a notice on the mailing package that the form is required to be returned by a specific date) or a compressed mailing schedule or both. In the experiment, three sampling strata will be used: high, medium, and low mail response areas. Each of the eight study panels will involve 10,000 households. These eight panels are: (1) control, (2) compressed mailing schedule panel, (3–5) three deadline messaging panels, and (6–8) three compressed schedule combined with deadline messaging panels. The three deadline messaging panels have language of varying degrees of sternness related to delays in mailing back the questionnaire. The analysis will focus on response rates, speed of response, and item nonresponse rates. Our only concern about this experiment is the lack of specification of the statistical power.

In summary, as the Census Bureau finalizes its preparations for the 2010 census, the panel believes that the Bureau faces tremendous risk if it does not perform comprehensive systems testing—focused on the interfaces between individual system components and, ideally, involving some field work component. The quality and utility of 2010 census evaluations will also be seriously impaired if census operational systems are not designed to retain procedural data for construction of a master trace database. The Census Bureau has proposed four experiments to be conducted during the 2010 census, but the panel believes that they suffer from design flaws and, significantly, lack connection to potential visions for the 2020 census. The panel suggests that three topics that are given little or no weight in the current CPEX plan—Internet data collection, use of administrative records in various census processes, and elicitation of accurate residence information—have greater potential to decrease the cost and increase the quality of the 2020 census, and so should be built into the 2010 experimental program.

We hope that the information and recommendations in this letter are useful to the Census Bureau. We would be happy to discuss and explain any of these issues at your convenience.

Sincerely,

Lawrence D. Brown, *Chair*
Panel on the Design of the 2010 Census
Program of Evaluations and Experiments (CPEX)

encl: Panel Roster

REFERENCES

Bentley, M., and J. G. Tancredo (2005). "Census 2000 Nonresponse Follow-up: Discrepancies in Enumerator Assigned Housing Unit Status," In *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

Hill, J. M., and J. D. Machowski (2003, September 29). *Master Trace Sample.* Census 2000 Evaluation B.6. Washington, DC: U.S. Census Bureau.

National Research Council (2004). *Reengineering the 2010 Census: Risks and Challenges.* Panel on Research on Future Census Methods, Daniel L. Cork, Michael L. Cohen, and Benjamin F King, eds. Committee on National Statistics. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council (2006). *Once, Only Once, and in the Right Place: Residence Rules in the Decennial Census.* Panel on Residence Rules in the Decennial Census, Daniel L. Cork and Paul R. Voss, eds. Committee on National Statistics. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council (2008). *Experimentation and Evaluation Plans for the 2010 Census: Interim Report.* Panel on the Design of the 2010 Census Program of Evaluations and Experiments, Lawrence D. Brown, Michael L. Cohen, and Daniel L. Cork, eds. Committee on National Statistics. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Tancredo, J. G., and M. Bentley (2005). "Determining the Effectiveness of Multiple Nonresponse Follow-up Contact Attempts on Response and Data Quality." In *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

West, K. K., J. G. Robinson, and M. Bentley (2005). "Did Proxy Respondents Cause Age Heaping in the Census 2000?" In *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

PANEL ON THE DESIGN OF THE 2010 CENSUS PROGRAM OF EVALUATIONS AND EXPERIMENTS

LAWRENCE D. BROWN (*Chair*), Department of Statistics, Wharton School of Business, University of Pennsylvania
RICHARD A. BERK, Department of Criminology, University of Pennsylvania
ERIC BRADLOW, Wharton School of Business, University of Pennsylvania
IVAN P. FELLEGI, Statistics Canada (retired), Ottawa
LINDA GAGE, California Department of Finance, Sacramento
VIJAY NAIR, Department of Statistics, University of Michigan
JESSE H. POORE, JR., Department of Computer Science, University of Tennessee, Knoxville
NORA CATE SCHAEFFER, Department of Sociology, University of Wisconsin–Madison
ALLEN L. SCHIRM, Mathematica Policy Research, Inc., Washington, DC
JUDITH SELTZER, Department of Sociology, University of California, Los Angeles
STANLEY SMITH, Department of Economics, University of Florida
JOHN H. THOMPSON, National Opinion Research Center, University of Chicago
ROGER TOURANGEAU, Survey Research Center, University of Michigan, and Joint Program in Survey Methodology, University of Maryland
KIRK WOLTER, National Opinion Research Center, University of Chicago

MICHAEL L. COHEN, *Co-Study Director*
DANIEL L. CORK, *Co-Study Director*
AGNES E. GASKIN, *Senior Program Assistant*
MEYER ZITTER, *Consultant*