National Institute on Aging
National Academy of Sciences

Expert Meeting on Cost-Effective Household Screening Methods for the Health and Retirement Study

The Keck Center of the National Academies
Washington, DC
November 19, 2012

MEETING SUMMARY

Rev. June 6, 2013

Executive Summary

The National Institute on Aging (NIA) commissioned the Committee on National Statistics to convene an expert meeting to advise the NIA on approaches to identifying eligible participants aged 51 to 56 years old in 2016 for the Health and Retirement Study (HRS). Traditional methods of screening households for the HRS have been expensive; identifying individuals from a narrow age band and recruiting statistically meaningful samples of Hispanic and black minorities are two key cost drivers during previous waves of screening. The expert meeting considered alternative screening methods that potentially could be less expensive without sacrificing a high quality, representative sample. To be timely for 2016, the HRS must submit its application to the NIA requesting funds to conduct the screening in early 2014.

In her welcome remarks, Connie Citro indicated that the goal of the meeting had been carefully worded. The meeting organizers preferred the term "cost-effective" to "cost-efficient" to signal researchers' concurrent interests in lowering costs and preserving high-quality samples. Cost-effective screening was recognized as a topic central to the HRS and to many other surveys. Researchers consistently have to grapple with issues of screening: finding people of interest and recruiting them to respond to surveys.

Meeting participants discussed a number of approaches that the HRS might consider:

- ***Pursue a long-term agenda to build legislative and regulatory support for sharing information across agencies.*** Data, such as the Census Bureau Master Address File (MAF), should be considered a national resource. Participants recognized that Title 13 restrictions create costly research inefficiencies and there are major costs, including duplication of effort, to information silos.
- ***Leverage commercial data offerings***. The 2010 Census Match Study revealed that commercial databases provide robust universe files. While not all commercial databases are equal in terms of quality, the study found that addresses attached to full names and date of birth (either MM/YY or MM/DD/YY) were likely accurate. Focusing on a few variables will yield the most accurate lists. Age would be a good variable for HRS to focus on.
- ***Adopt National Longitudinal Study of Youth (NLSY) cohorts.*** If the HRS were to adopt the older cohorts from the NLSY, it would contain a comprehensive panel with a wealth of data for researchers to study. For instance, early cognitive data would be enormously

valuable for understanding the evolution of cognition and onset of dementia. One issue with this panel, however, is the lack of recent immigrants. The Office of Management and Budget (OMB) also would need to authorize the transfer of any NLSY cohorts.

- ***Adjust the target population to 46 to 56 year olds and screen every 10 years.*** The 2016 cohort would enroll 51 to 56 year olds. Then each year, individuals who turned 51 would be pulled into the survey from the pool of 46 to 50 year olds identified in the screen. This would transition the HRS into a steady state of recruitment.
- ***Employ multiple modes of recruitment to drive efficiency.*** Exhausting cheaper modes of contact (i.e., mail and telephone) prior to in-person contact would permit some savings.
- ***Pool resources with other surveys to screen a larger national sample frame and obtain a household roster with demographic characteristics.*** Demand from a constellation of survey research organizations and Federal agencies conceivably could support such a broad survey screen. Achieving long-term cooperation among survey research organizations to agree on specifics and timing will be challenging, especially given the limited shelf life of the data.
- ***Conduct further research on screening methods with NIA funding.*** In particular, the topics below were mentioned as of interest:
    - Understanding the costs associated with each step of recruitment including screening via mail, telephone, and in-person, including the biggest cost drivers.
    - Understanding the cost and effectiveness of tracing or tracking households with 46 to 50 year olds, so that these households could enter the HRS at a later date.
    - How incentives affect interviewers and respondents, including how increasing interviewer knowledge about eligibility and response rates affect recruitment rates.
    - Whether response rate mapping can improve efficiency. The Census Bureau is building a database mapping response rates to the MAF in order to facilitate research and testing for the 2020 Census.
    - The extent to which screening using the network sampling could be employed, its practical application, and its likely biases.

Screening Challenges Faced by the Health and Retirement Study

Chair: John Thompson, National Opinion Research Center (NORC)

Background on the Health and Retirement Study and Rationale for the Meeting

John Phillips, National Institute on Aging

The mission of the NIA is to improve the health and wellbeing of older Americans through research. The NIA's Division of Behavioral and Social Research (BSR) is charged in part with administering infrastructure grants that support research including longitudinal aging studies, such as the HRS.

The HRS is a significant data infrastructure project that has proven to be a useful resource for public and private researchers, serving as the basis for more than 1,700 publications as of November 22, 2010. It makes publicly available a nationally representative, population-based sample of more than 20,000 Americans over age 50 from 1992 to 2010. Administrative, clinical, and genetic data complement the rich survey data collected in the HRS. The data are used by health researchers representing a wide array of disciplines, as well as US federal agencies including: the Office of the Assistant Secretary for Planning and Evaluation (ASPE), Congressional Budget Office (CBO), OMB, Social Security Administration (SSA), Treasury, Federal Reserve Board, and Government Accountability Office (GAO). The University of Michigan manages the HRS through a grant that is funded as a cooperative agreement from the NIA, with cofunding from the SSA.

In order to maintain the national representativeness of Americans over age 50, HRS must add new participants periodically. Though it started as a single cohort study in 1992, the HRS moved to a steady-state design by 1998. Different strategies can accomplish steady state. The HRS employed a strategy of adding targeted birth cohorts aged 51-56 every six years; to date, it has added six cohorts to ensure age representation and in 2010 expanded the minority sample.

HRS plans to add a new cohort in 2016 of individuals born between 1960 and 1966. Researchers expect that the screening and recruitment will be costly, as it has been in the past. Therefore, the NIA has brought together this group of experts to answer two questions: 1) Are there alternative, more efficient screening methods that are high quality and less expensive? 2) How might the NIA/HRS approach screening in the short run (2016) and long run?

Richard Suzman added that an efficient screening approach could have a major impact on many other studies. He offered three partial solutions for the group's consideration. The first option is to consider if the HRS could sample from SSA records, which likely would require adoption of several changes to current legislation. The second option is to design an exchange or auction system, whereby the University of Michigan could permit access to parts of the sample identified through screening as ineligible for the HRS but of value to another study, for example, individuals below age 40. It could be a particularly useful strategy if there were other researchers

with a narrow age interest.[1] Finally, a third option would be to approach the NLSY researchers about taking on their aging sample.

Sampling Rare Populations

Graham Kalton, Westat

Graham Kalton discussed the challenges of sampling rare populations and the advantages and disadvantages of methods that might provide efficiencies. Ultimately, he concluded that there is no "magic bullet." He cautioned against the use of inexpensive screening methods that would result in greater noncoverage and/or nonresponse. He emphasized that benefits of a high-quality sample would accrue over time and that comparisons across different approaches be based on net present values that capture both costs and benefits.

Kalton summarized his understanding that the HRS is planning to add 4,000 persons aged 51 to 56 in 2016. According to October 2012 Current Population Survey (CPS) data, only 18 percent of households have at least one resident between 51 and 56 years of age, suggesting that over 22,000 households would need to be screened. Could a more efficient sample design generate a high-quality probability sample of this age cohort? Would oversampling of minorities be required? Would the HRS sample need to be clustered?

A few points were clarified regarding the needs of the HRS. Some screens would be conducted by phone. Clustering would be required for field interviews. David Weir explained that strictly speaking, the HRS sample consists of financial units, not individuals.

Several probability sampling techniques for rare populations are available in addition to large-scale screening, including disproportionate stratified sampling, network (multiplicity) sampling, multiple frames (i.e., multipurpose screen), two-phase sampling, location sampling, and accumulating or retaining samples over time.[2] Selecting the best screening method requires consideration of two key issues, as related to the target population of interest (as distinct from the general population): response rates and coverage rates. Response rates to screening and the main survey may be adversely affected by survey methods. Low coverage rates may arise in the original sampling frame or due to false negatives.

Kalton discussed the potential use of a host survey for screening. The CPS and the American Community Survey (ACS) are examples of large representative samples that could potentially serve as hosts (although Title 13 restrictions would be a major obstacle). In general, many issues arise when using a host survey for screening: survey managers and institutional review boards (IRBs) must agree; data collection must be compatible; the host survey must be high-quality; and informed consent for multipurpose screening must be obtained. As a result of these issues, although there are some examples of this approach, the use of a host survey is uncommon. Citro

[1] As example, the National Social Life, Health, and Aging Project (NSHAP) sample of community-residing adults born between 1920 and 1947 derived from the national household screening carried out by the HRS in 2004.

[2] See Kalton G. 2009. Methods for oversampling rare subpopulations in social surveys. *Survey Methodology* 35 (2): 125-141 (December).

added that not only would individuals need to provide consent, but the OMB would need to review and approve the use of a government survey screen for other purposes.

Similar to using a larger host survey for screening, the HRS could attempt to construct a multipurpose screening survey. Researchers studying different rare populations could pool resources to conduct a large-scale screening survey, preferably with populations that do not overlap (i.e., disjoint populations). It could be useful if such a survey was able to provide smooth transitions to the main survey data collections for each of the rare populations.

A number of efforts were mentioned as examples of multipurpose screenings. The National Center for Health Statistics (NCHS) uses the State and Local Area Integrated Telephone Survey (SLAITS) mechanism as the basis for a number of health surveys, including the National Immunization Survey (NIS) that is conducted by the NORC at the University of Chicago for the NCHS. At one point the National Health Interview Survey (NHIS) was conceived as an umbrella survey that could support multiple individual spinoff surveys. One round of the National Survey of Family Growth was based on retired NHIS cases. The sample for the Household Component of the Medical Expenditure Panel Survey (MEPS) is selected each year from previous year NHIS respondents. The data collected in the NHIS is used for oversampling certain households (e.g., by race/ethnicity, and other groups of topical interest) for MEPS. The design also provides the potential for linking the NHIS and MEPS responses for analyses. The NHIS is a 90,000-person sample that would be large enough to support the HRS need for a replenishment sample. Suzman raised potential issues with the use of the NHIS for this purpose with respect to its oversampling of minorities and its widespread sample across many primary sampling units (PSUs).

Suzman supported the idea of a multipurpose screen, suggesting that statistical firms could share in the upfront costs, and could then market and sell portions of the sample to certain government agencies or other researchers. Weir pointed out that private companies would be wary of taking on the financial risk of running a large screen to develop the sample, without first knowing that customers were in place.

Screening economies can be obtained by collecting screening data about all members of a sampled household from a single household informant. Further economies can be obtained by the use of network sampling in which sampled individuals report their own screening status and also the statuses of those living elsewhere who are linked to them in a defined way, such as siblings, parents, and children. There are a number of practical limitations that have limited the widespread use of network sampling. Another technique for extending the screening data collection more widely is known as focused enumeration. With this sampling technique, household informants report the screening data for everyone in their own household and also for their neighbors' households (e.g., two households on either side of their households). This design has been used with the British Crime Survey. A potential limitation to focused enumeration is it may lead to serious undercoverage of the population that is to be oversampled. The technique appears unlikely to work well for the HRS because people are generally unlikely to be able to accurately identify neighbors aged 51 to 56, a narrow age range. In order to reduce the amount of false negatives, Kalton suggested adopting an expanded age range (e.g., persons aged 45 to 60) for the initial screening.

Disproportionate stratification, oversampling strata where the rare population is more prevalent, can also be used to make screening more efficient. However, there does not appear to be an application of this method that would be useful in screening for the 51 to 56 year old age group for the HRS. Geographically based disproportionate stratification is often effective for oversampling of some racial groups and for Hispanics. When the sampling frame contains them, names can be used to form strata with differing prevalence of certain racial groups; for instance, the California Health Interview Survey used this approach to oversample Koreans and Vietnamese.

Disproportionate stratification works well when three conditions are met: 1) the rare population must be much more prevalent in the oversampled strata; 2) the oversampled strata must contain a high proportion of members of the rare population; and 3) the cost of a screening interview must be low relative to the per unit cost of the main data collection. In the case of the HRS, it seems the third criterion is fulfilled (lower screening cost than main data collection cost), but there do not appear to be ways to construct strata such that the first two criteria are met. Kalton offered a formula to calculate the optimal relative sampling rate based on the combination of these three criteria. An efficient disproportionate design takes all three of these criteria into account. He pointed out that oversampling a stratum with a high prevalence at a very high rate in order to produce a large sample size for the rare population is likely to be a very inefficient design; the wide variability in the sample weights would lead to a great loss in precision for the survey estimates. When disproportionate stratification is used, the sample size objective should be expressed in terms of the effective sample size not just simply the sample size. The determination of the optimum allocation across strata includes the relative cost of the main data collection to the cost of screening out ineligible participants. This point led to a discussion of the need to better understand the costs of each recruitment step, including training and interviewing.

Kalton briefly discussed address-based sampling (ABS) as a more cost-efficient method for area sampling than conventional listing. With ABS, surveyors can buy samples of United States Postal Service (USPS) residential addresses from licensed vendors, who can also enhance the addresses with additional information such as telephone numbers and demographic information. While ABS can reduce sampling costs compared to area sampling for in-person surveys, survey procedures need to be put in place to address noncoverage in the USPS lists. For example, some housing units have only PO box addresses and others have rural route addresses, neither of which can be readily identified by the survey interviewers. Procedures also need to be put in place to deal with drop points, that is, single addresses for the delivery of mail that contain multiple households, such as large apartment buildings in cities.[3]

Three take away points concluded the presentation on rare population sampling. First, there is no magic bullet for cost-effective screening. Second, benefits from a high-quality sample will accrue from all subsequent waves of the HRS. Third, economies in screening that result in greater non-coverage and/or nonresponse are likely to prove counterproductive in the long run.

[3] Mansour Fahimi's presentation covered ABS more extensively in a later session.

Sampling Challenges and Opportunities Specific to the US Federal Statistical System

Robert Groves, Georgetown University

The sampling challenges that the HRS faces are a subset of all US sampling issues. The Federal statistical system features large, high-response rate household surveys that can be attractive to other survey organizations as a potential sampling frame. While there is some oversampling of minorities, such sampling is not ubiquitous. Some agencies use sampling frames based on the MAF. Dispersed agencies are relatively free to choose sampling frames, modes of data collection, etc. Myriad contracts have been issued over decades to fund surveys that have an important statistical component for the government, but the contracting has never been coordinated. All Federal surveys share confidentiality protections under the Confidential Information Protection and Statistical Efficiency Act (CIPSEA).

Potential legislation that would allow more data sharing has moved at a glacial pace. As an example, it took 10 years for the first legislation affecting data sharing to pass. It is likely that further enabling legislation has not passed because of a lack of urgency. NAS made recommendations in support of new legislation; however, that did not impact or change legislative activity. Lack of supporting legislation, therefore, must be a factor when considering options available to the HRS moving forward.

There are several familiar opportunities for the HRS to more efficiently screen new cohorts. The first option is to re-interview age-eligible persons in households measured in other surveys. The second is to use screened out households from other surveys by locating a demographic subset of the sampled population. The third is direct use of the MAF.

Re-interviewing age-eligible persons has several advantages and disadvantages. On the one hand, re-interviewing would provide better estimates regarding likelihood of success; permit "free" acquisition of prior wave reported attributes; and lower cost when compared to novel screening. On the other hand, this method has several drawbacks. Processing lags from the first survey often lead to nonresponse in subsequent waves because respondents have moved. When the NHIS became the first sampling frame for MEPS, the first few years were inefficient for this reason; sample cases were not available for second case interviews quickly enough. A potential cost driver for this method is tracking individuals who move. Another hurdle is the nonstandard nature of informed consent across different surveys that use different IRBs. Researchers must consider IRB requirements and potentially OMB review. Sufficient numbers of age-eligible HRS cases are not readily available.

Use of screened-out households from other surveys has similar advantages and disadvantages. Among the advantages, there are reduced costs of screening, some surveys supply household roster data (i.e., information about all occupants of the household), and it is possible to estimate propensity to respond given paradata[4] from the previous survey. Issues with this approach include: measurement error on screening variables; processing lags that often cause nonresponse

[4] Paradata refers to data about the process of collecting survey data, such as interviewer call records, length of interview, keystroke data, interviewer characteristics, interviewer observations, mode of communication, etc.

from failure to locate; nonstandard informed consent regarding re-contact; and insufficient sample size.

Direct use of the MAF was an active topic of discussion. The MAF is desirable because "fresh" sample cases are uncontaminated by recent survey contact and frame development costs are reduced. However, direct use of the MAF involves known legal issues, as well as other drawbacks. The Supreme Court ruled that the MAF is covered under Title 13. A legislative change is required to change the legal basis for using the MAF. Additionally, the Census Bureau has concerns over unanticipated impacts of using the MAF. In the 2010 Census, there was no language that suggested to participants that someone else might re-contact them for a follow-up survey. Even if legal issues were resolved, a mechanism for sharing the data still would need to be appended. Mansour Fahimi suggested using the MAF as a supplemental list, rather than as a cornerstone, due to its lack of demographic data.

Groves concluded with thoughts on how to move forward. He reiterated the (Title 13) legal prohibitions governing use of the MAF. The ACS could be a potential sampling frame; though, it is not clear what would happen if there were follow on surveys or what the contract would look like. Minority groups' concerns over information about illegal residency would need to be alleviated, likely by not sharing those data outside of the Census Bureau. Groves proposed sharing a MAF of structure designations, not a MAF of housing units. Another approach would be to seek direct permission from the Census Bureau survey respondent for a nongovernmental agency (Federal contractor) to contact them for a follow-up survey.

Citro acknowledged that developing the MAF into a public utility is doubtless an attractive proposition. But given the known legislative barriers, she asked how much effort should be allocated toward pursuing this option? Overall, participants seemed to agree that legislative change is worthwhile to pursue given the potential of the MAF as a national resource. However, such change was seen as a long-term proposition, thus interim efforts must focus on accessing other lists that do not pose these fundamental issues.

The group considered precedents outside the US government making MAFs available for research purposes. Austria is the only country in Europe known to offer such a resource. In Sweden, the government was just building its address registration revenue agency when its citizens began to protest. The American public probably would be even more reluctant to allow this to happen.

The SSA mailing list seems like another source of a universal address list. Howard Iams pointed out that there is still a legal barrier: Title 26 of the US Code allows access to addresses for the sole purpose of mailing Social Security statements. Even if that were removed, SSA information is about individuals, not the household unit that is the sampling unit for the HRS. Additionally, SSA has poor data on minority status. The extent to which the SSA mailing list reflected the 2010 Census was discussed further in Amy O'Hara's presentation.

John Phillips asked if a multipurpose screen might be a better line of approach given that historical attempts at changing Title 13 have not succeeded. Groves considered success dependent on a stable consortium of agencies willing to support a shared sampling strategy and

to maintain the effort over time. He observed that effective coordination would require abundant goodwill from participants. In order for this to work, all potential users must be at the table early, before contracts are in place because changes post hoc are so cumbersome. The next step would be to consider specific surveys to ensure minimal overlap of target populations. Minority samples from surveys such as the ACS are quickly consumed.

The ACS was discussed further as a sampling frame. Logistically, if the ACS was used, a formal request would have to be made and the OMB would coordinate competing demands. Four members of the Federal Interagency Council on Statistical Policy would act on such requests. Guidance for making a request is available on the Census Bureau and OMB websites. Beyond the burden of getting access to the ACS sampling frame, it is not clear that the ACS provides the best list for the HRS. The ACS collects continuous monthly data and weights results over the year. Data files are available the following year in August. By the time a survey such as the HRS could field a screener, the ACS data could be two years old, raising a relevant time frame issue.

Underreporting of Populations in Screening Surveys

Frauke Kreuter, University of Maryland

Kreuter described research findings from a joint project with Roger Tourangeau and Stephanie Eckman that may be used to mitigate underreporting of populations in HRS screening. Her research led to several conclusions: both interviewers and respondents are sources of underreporting; incentives for more accurate reporting include strong manipulations of salience and burden/reward; and a major caveat is a tradeoff between participation and motivated underreporting.

Kreuter reported from an integrated set of studies designed to examine three potentially related forms of measurement error. These errors may have large effects on the results from surveys. One form of error involves underreporting in response to screening questions. Many surveys begin with a short "screening" interview that is designed to identify members of the eligible population. Here underreporting can lead to serious coverage errors. The second form of measurement error involves underreporting in response to filter questions. Most surveys attempt to route respondents around questions that don't apply to them; respondents aware of such survey structure might give false negative responses to the filter questions to avoid follow-up items. The third type of error is similar: if respondents remember the survey structure in a repeated administration, they may answer differently to avoid a lengthy part of the survey. Researchers generally hypothesized that the desire to reduce effort needed to complete the survey led to errors.

Researchers at the University of Maryland College Park (UMCP), Westat, and the Institute for Employment Research in Germany, posed a question about the source of these errors: do the effects reflect respondents' motivation; interviewers' motivation; some combination of the two, or are the effects instead a communication misunderstanding? Furthermore, could errors be reduced through changes in salience (i.e., how obvious is it that one can save time by changing an answer?) or burden/incentives (what is the cost associated with underreporting)? To study these questions, Kreuter and Tourangeau secured funding from the National Science Foundation (NSF) and the Institute for Employment Research in Germany to conduct a series of experiments to manipulate the error mechanisms in both the screening and the questionnaire.

In the screener experiments, researchers studied the impact of salience by sending an advance letter describing the survey and eligibility and changing eligibility question wording from open to direct. Interviewer payment was manipulated to study impact of incentives.

Researchers found that revealing the target population changes eligibility rates. Advance letters had a mixed effect that seemed to fade over time. Higher eligibility rates were attained using a household roster format. Though there was a tradeoff between coverage and nonresponse error: full roster increased coverage, but response rates were also lower. It was noted that coverage rate and nonresponse rate had to be considered together to attain an accurate interpretation of which methods would lead to the most effective screening.[5] Weir noted that the HRS used a roster screen in the past and experienced abysmal response rates.

The discussion about household rosters was extended to discuss network sampling, specifically asking informants about their own household and neighboring households on each side. Using this method, informants provide imperfect information about their own households (e.g., age), and provide less perfect data about their neighbors. It would be ideal to find the right balance between high-volume and high-quality of data. To find that balance, cost data are needed including the cost of false negatives.

Additionally, researchers need to understand how network-sampling data compare with commercial lists. Richard Valliant described some of his group's experience with MSG's enhanced address file. They found that among housing units where MSG had no complementary data, 25 percent were vacant. Information about vacancies and misclassifications would improve sampling efficiency. Fahimi indicated that vacancy and misclassification rates are available and that generally error rates go up as more variables are appended to enhance address files.

Finally, there was discussion about how other mechanisms may affect reporting and coverage rate. In the studies that Kreuter discussed, interviewers knew some age data based on voter registration that was more likely to produce false positives than false negatives. Tourangeau stated that interviewers may have been more likely to confirm information they had than vice versa.

Methodological Advances and Opportunities that Could Benefit the Health and Retirement Study

Chair: Roger Tourangeau, Westat

Ongoing Research on Alternative Screening Methods for the Health and Retirement Study

Richard Valliant, University of Maryland

HRS goals have changed over time. The 2004 HRS sample cohort included Early Baby Boomers (EBB) born from 1948 to 1953 and screened for Middle Baby Boomers (MBB) born from 1954

[5] See Tourangeau R, Kreuter F, and Eckman S. 2012. Motivated underreporting in screening interviews. *Public Opinion Quarterly* 76 (3): 453-469 (Fall). First published online: August 31, 2012. http://poq.oxfordjournals.org/content/early/2012/08/28/poq.nfs033.short.

to 1959 for later use. Each age cohort had about 4,000 participants. In 2010-11 the HRS recruited MBBs, oversampling Blacks and Hispanics. EBBs and MBBs are each about 8 percent of the population. EBB Hispanics and MBB Hispanics are about 1.5 percent of the population. Any HRS sample design must account for rareness of target groups in order to reach efficiently desired sample sizes.

The HRS sample design is comprised of 75 primary sampling units (PSUs)—counties or groups of counties. Large counties such as Chicago, New York City and Los Angeles are included. Additionally, there is a probability sample to represent the whole country. About 840 segments—combinations of Census blocks—made up the 2004, 2010, and 2011 cohorts combined; households are sampled from these segments. In 2004, 2010, and 2011, about 50,000 household units cooperated with screening. Over one in five households had one or more age-eligible persons; the average number per screened household was 0.37 in 2004, 2010, and 2011 combined. For the last two screens (in 2010 and 2011), there were only 0.30 eligible persons per screened household. At a gross level, increasing the number of average eligible respondents per household would significantly lower cost.

Several sampling changes took place along the way. Midway through 2004, the HRS began to stratify segments into four groups. These strata were used to identify areas that had higher concentrations of Black and Hispanics. Strata were defined as follows: 1) Blacks and Hispanics each comprising less than 10 percent of the population; 2) Less than 10 percent Hispanic; more than 10 percent Blacks; 3) Less than 10 percent Blacks; more than 10 percent Hispanic; and 4) Blacks and Hispanics each comprising more than 10 percent of the population. Stratum 1 contained 65 percent of all segments; strata 2 and 3 each had 15 percent of segments; and the remaining stratum 4 had 5 percent of segments.

MSG provided data on addresses for three of the four strata that included populations with more than 10 percent of either or both minorities. It evaluated the impact of using out-of-date data (2000 Census) to stratify segments in 2004, 2010, and 2011, and found that out-of-date data did not have a major adverse effect. MSG used overall rates for sampling persons in each of the segment strata. Researchers compared the expected number eligible if segments were stratified using old Census 2000 data to the expected number if segments were stratified using up-to-date data.

Another evaluation looked at optimal allocations of segments (i.e., different ways to stratify the sample) compared to an equal probability sample (i.e., roughly screening 10 households to find one eligible household). Researchers attempted to account for the costs of screening and interviewing, proportion of eligible persons in each stratum, and proportion of domain populations in each stratum. For nonminority age-eligible strata, very little gain was observed from optimal allocation. For strata with more than 10 percent Black and/or Hispanics, there were gains that persisted over the decade between Censuses; however, those gains decreased as interviewing costs increased. In the HRS, the time to complete one interview costs about the same as three completed screens. Valliant concluded that geographic stratification with an optimal allocation is beneficial even if stratification data are out-of-date but only for the rarest domains. Optimal allocation also can be useful for hitting target sample sizes in domains.

Commercial lists can be used for improved sampling. These lists are based on the USPS Delivery Sequence Files (DSFs). Demographic data are added to address listings. Lists can be used to replace or augment field listings or to stratify addresses for sampling. Addresses from segments with more than 10 percent Hispanics or Blacks were sent to MSG. Substrata of addresses were formed based on MSG information about whether any household occupants were MBB or EBB and whether they were Hispanic or Non-Hispanic. A sixth category captures unknowns, of which HRS fieldwork found 25 percent of household addresses to be vacant. Table 1 summarizes the comparison of MSG data to HRS fieldwork.

Table 1. Summary of Classification Results using MSG Data

MSG classification	Unweighted, percent of households (n=28,164)	Weighted, percent of households (n=38,943,822)
Correctly classified	29	40
Incorrectly classified	21	22
Unknown	50	38
TOTAL	**100**	**100**

One method for calculating stratum-sampling rates uses a linear program. By finding the sample number of households in each stratum, the number of households that must be screened is minimized. This method is subject to a couple of constraints. First, the expected number of persons in each domain must be greater than or equal to the target number eligible in the domain. Second, each stratum sample size must be greater than or equal to the minimum.

The average number of eligible persons per household based on what HRS screening found was compared to the MSG classification frame for strata 2-6. Notably, it was acknowledged that MSG data on stratum 1 likely were more accurate. The MSG classification was correct a majority of the time ranging from 56 to 78 percent accuracy. The HRS screening picked up other eligibles that may have been classified incorrectly (e.g., MSG classified the household as non-Hispanic MBB; the HRS team identified a Hispanic EBB in the household). The fifth and sixth strata that MSG classified as having no eligibles or were unknown yielded very few positive HRS screens. Regardless, the HRS screens these likely ineligible households because it wants a probability sample of the entire population.

Weir noted that screening these likely ineligible households affects the sample weights: one out of 15 households that screen in will come from strata 5 and 6. Furthermore, it is likely that the households that fall into these two strata have different characteristics than other households. For instance, unknowns are less likely to have credit cards and are economically different than eligibles in the first four strata. Assumptions that would allow weighting variation to go away are, therefore, likely unjustified. The group acknowledged that the HRS must be a representative sample, using statistically sound sampling methods that would receive a high mark from peer review. Any cost cutting method should not jeopardize sample representativeness.

Based on the above analysis, the linear program method would only sample 200 of the likely ineligible households because the yield is so low. However, the constraints on the final sample size also affect the weights. The HRS sought a 9,000-household sample with two cohorts of

around 4,500. One way to reduce the screening effort is to improve effective sample size. The average number of eligible persons per sample household in the linear program solution would be 0.82. This compares to the average of 0.37 across the three cohorts (2004, 2010, and 2011), and the average of 0.30 for 2010-11. Sampling so few households in strata 5 and 6 may be too extreme.

Planning for the 2016 HRS recruitment is just beginning. The team expects to take certain steps, including: 1) updating the sample of geographic areas (PSU) to account for Census 2010 data; 2) redefining segments (groups of census blocks); 3) purchasing a commercial list of addresses with demographic information; and 4) extending analysis to apply to the national sample. Researchers agree that it would be good to cut costs; though at this point it is difficult to estimate how sizeable a cut is achievable.

Weir initiated discussion about minority sampling and tracking of future eligibles. If the HRS wanted to replicate a 4 to 1 oversample of minorities, then the HRS would have to do a 2 to 1 screening sample. If the 2010 screen identified minorities who would be eligible in 2016, then tracking those eligibles could be productive. Historically, the HRS has had difficulty tracking people and has attempted to obtain household roster information that would help with tracking at a later point. The NLSY successfully tracked respondents. While the HRS main interview provided information that improved tracking, the screener gathered limited information about tracking people who would be eligible six years later.

Geographic data could potentially improve screening efficiency. Ethnic groups tend to cluster in geographic areas. However, the elderly tend to be dispersed. Because improving efficiency relies on identifying highly-concentrated groups of the target population, it is unlikely to prove fruitful for HRS screening. Unless the SSA could release individual data, the HRS only could analyze age distribution at the county level, which would be too coarse. Only average age data would be available. In response to a question from Suzman about whether the Census can provide data on the density of specific age groups in a geographic area, Valliant noted that the HRS can form sample blocks based on ACS data. However, Kalton indicated that the ACS provides only 5-year average estimates at the block level, and furthermore these estimates are very imprecise. The ACS is post-stratified at the county level to the census or population estimates. Fahimi suggested use of Nielsen-Claritas data as an option, although the data are proprietary and thus not very transparent; Valliant noted that the National Household Survey on Drug Abuse used Nielsen-Claritas for small area modeling.

The discussion concluded by recognizing that weighting would need to be considered further, particularly for strata with smaller eligible populations. Trimming weights only could be applied if there were sound methodological justification.

Advances in List-Based Sampling Methods

Mansour Fahimi, Marketing Systems Group

Fahimi presented an overview of MSG advances in list-based sampling methods. Because proportional sample allocation is not suitable for domain analysis and is expensive for rare domains, he recommended disproportionate sample allocation. Disproportionate samples can be formed using ABS or dual-frame random-digit dialing (RDD). List-assisted ABS and RDD are

both suitable for the HRS and can reduce recruitment costs, increase sample size, and reduce survey administration costs.

As background, the HRS surveys people transitioning from workforce to retirement. It looks at their financial situations, social activities, and health status. Thus far, the HRS has incorporated three sample cohorts in 1992, 2004, and 2010. The HRS costs $15 million to recruit and survey 4,000 individuals; screening out ineligibles represents two thirds of the cost. One factor that drives cost is the fairly low incidence of eligible households, about 13-22 percent of all households according to March 2012 CPS data. Alternative sampling methods could reduce these costs.

Increasingly, interest has focused on ABS. Several reasons support growing popularity of this sampling alternative. Dual frame methods have led to ad hoc sample mixtures and inconsistent weighting applications. Over time, landline RDD coverage is declining because of cell phone only (CPO) households. A less known issue is that landline RDD now pops up outside the traditional frame (e.g., VOIP accounts). Response rates to single modes of contact also have eroded, a trend that is not unique to geography, mode, or sponsor, and that is associated with increasing costs of refusal conversion strategies. In contrast to these challenges, improvements of household address databases (e.g., DSFs) are enabling more efficient sampling.

The latest version of the DSF contains 139 million addresses. Most are city-style residential addresses (116 million). The remaining 23 million addresses are for the most part represented by traditional Post Office boxes (14 million), vacancies (4 million), and drop units[6] (2 million). Not all vacancies are truly vacant. Thompson noted that the Census estimated a vacancy rate of around 8 to 9 percent, higher than the data presented.

There are potential issues with using the DSF for ABS. These issues and potential solutions are summarized in Table 2.

Table 2. Potential Issues and Solutions for ABS via DSF

Problem	Solution
Incomplete rural addresses	Wait for completion of 911-address conversion
Frame multiplicity (i.e., duplicate households)	Remove regular Post Office boxes that correspond to other addresses; remove vacant only way of getting mail (OWGM) boxes
No contact information	Add names and phone numbers from commercial sources
No geo-demographic data	Geocode addresses to Census blocks and apply CPS and ACS demographic profile data; add household demographics from commercial sources

Coverage problems in rural areas because of incomplete addresses are a diminishing issue over time because 911-address conversion has simplified addresses. MSG mitigates the problem of frame multiplicity (that causes duplication of effort) by removing regular Post Office boxes that

[6] Drop units describe addresses where mail is dropped for multiple households and then delivered internally.

correspond to other addresses and removing OWGM boxes. More importantly, the DSF by itself lacks geo-demographic indicators that are needed for sample stratification. Geo-coding of each address to a Census block using Census geographic definitions and CPS and ACS demographic profiles can minimize this issue. A final challenge is the lack of contact information. MSG can append names for almost 90 percent of addresses and names and telephone numbers for more than 50 percent of addresses. Geo-demographic information can be added for pennies per address. Appending commercially available ancillary data can enhance ABS further: person-level information (e.g., age, race, education), household-level information (e.g., size, income, purchases), and inferred information for household and area. Commercial data sources obtain data from credit reporting agencies, warranty cards, etc.

Dual frame RDD is an effective sampling methodology for certain surveys. There are about 10 million 100-series landline telephone banks. A telephone bank is a consecutive sequence of telephone numbers. Of the 10 million, only about 3 million are included in the list-assisted RDD frame. 4.3 million landlines do not have a listed bank number but they do at the exchange level (Genesys). 2.6 million are not listed in banks or exchanges. MSG is increasing efficiency of its sample; there are more and more landline numbers in these banks.

Trends in households with landlines tell an interesting story. Between 2005 and 2010, listed landline numbers declined from 77 million to 62 million. The percentage of households with landlines steadily increased between 1963 to its peak in 2001, with 97 percent of all households owning at least one landline. That trend reversed quickly starting in 2003. A decade later, only 66 percent of all households had a landline. CPO households grew from 4 percent in 2003 to 32 percent in 2011. Targus provides these data to MSG.

Pew Research on a landline sample found that in the heyday of landlines, age composition was fairly well distributed. In 2010, the distribution was heavily skewed toward the over age 50 population with this group representing 67 percent of landline households. Individuals aged 18 to 29 represented only 7 percent of landlines. Eighteen to 36 year olds largely represent CPO households; though a significant proportion of CPO households have a member over 50 years old.

The HRS should consider a few potential issues with future use of dual frame RDD. Currently, there is a fair amount of inconsistency between designing the RDD sample and the backend weighting. The inconsistencies seem to rely on dual frame RDD from recent election polling. For CPO households, MSG must rely on data from the Centers for Disease Control and Prevention (CDC) that are based on surveys prone to error. The data are reported at the state level, limiting usefulness. These data do not capture variation within states at the county and regional level.

Dual frame RDD could be refined. MSG can improve the landline frame with exchange level demographic information; though age is not necessarily one of those. Ancillary data from commercial sources for listed numbers would help as well. For the cellular frame, samples can target geographies via rate center boundaries. To a limited extent, samples can target certain demographic populations via rate center profiles. Certain billing zip codes can be appended post-sampling as well.

MSG can provide county-level CPO estimates for every county in the United States (approximately 3,700). Thompson suggested that overlaying county-level CPO data with population data would provide additional insights. MSG develops its estimates starting with counties and the total number of telephone households from ACS. Then it determines whether a telephone number is associated with a landline household or a CPO household, or a household with both. From a separate source, they know about how many landline households there are. The difference between the two yields the number of CPO households. Use of ACS data could be problematic because the ACS calculates 5-year averages.

Surveys often rely on disproportionate stratified sampling to accommodate analytical objectives and reduce screening costs. When allocating sample to strata three factors should be considered: stratum-level cost, stratum-level variability, and unequal weighting effect. Complex optimization requires more assumptions. Using a formula, MSG calculates the effect of disproportionate stratified sampling. Any clustering can be calculated. Along with the weighting, it obtains a rough metric on a core weighting effect. An example comparing proportional and alternative allocations illustrates that oversampling from the productive sample leads to a reduction in effective sample because of the design effect. This observation leads to the question: How does one oversample, and reduce cost, without lowering sample quality?

The mathematically relevant answer to the above question is the intersection of two lines. The first line plots the effect of unequal weighting by percent of sample allocated to the productive stratum (i.e., design effect). The second line plots the number of screened households to recruit 4,000 participants by percent of sample allocated to the productive stratum (i.e., cost). This illustrates the inverse relationship between more inaccuracy and lower cost (i.e., quality and cost are directly related). Thompson added that the math is more complicated depending on the study design. He mentioned preliminary findings from an unpublished paper by NORC researchers that suggest that the same accuracy can be achieved while reducing costs by 15 percent using age-targeted lists from the NHIS for the NIS. The NIS also has a complicated design involving rare variables. In both the HRS and NIS scenarios, it is essential to understand the exact cost structure of screening and interviewing.

Discussion centered on the quality, depth, and breadth of MSG data. It was unclear whether MSG datasets have good coverage of poor Americans. Factors affecting quality include level of geography, types of addresses, and number of variables. If the database were only enriched with age, accuracy would appreciably improve. Completeness of household-level data was also unclear. In 10.8 percent of households in one of MSG's databases, there was at least one person age 50 to 55. Better data on whether or not more than one eligible individual is present in a household would be helpful.

Regarding MSG data compilation methodology, Valliant inquired about processes in place to resolve data conflicts between databases. Fahimi clarified that MSG has in place a hierarchy of decision rules about which data to use to resolve data conflicts. For example, priority may be given to US Info or Targus when it comes to age.

State-Based Address Lists

Amy O'Hara, Census Bureau

A state-based address list can come from two sources: 1) independent data collection, as with the 2010 Census, or 2) third party provider. Vendors offer their commercial data at point in time, historical, state-level, and/or substate-level. The Census Bureau conducted a study to assess the value of commercial databases for researchers and found that indeed commercial data cover the universe in a robust way.

The Census Bureau issued a Request for Proposals (RFP) for the MSG report and acquired person and address universe files to evaluate their utility. Five vendors—Experian, Targus, VSGI, Info USA, and Melissa Data Base Source—provided a total of nine datasets. These datasets contained various individual-level data including: address, email, race, age, gender, date of birth, social security numbers (SSNs), utility records, cell phone records, etc. The vendors agreed to licensing terms subject to the Freedom of Information Act (FOIA), while the Census Bureau agreed not to release results that compared specific vendors. All vendors that responded to the RFP were reluctant to share information about their methods of data collection, data sources, and data cleaning processes. Importantly, the Census Bureau had to trust that vendors sourced information properly and obtained consents, particularly two lists that provided SSNs.

These datasets were compared to the MAF and Federal data from the Internal Revenue Service (IRS), Selective Service, Medicare, etc. With data use agreements already in place, the Census Bureau has access to a vast amount of data. After de-duplicating all the Federal and commercial data in the Statistical Records System (STARS), the Census Bureau sought to understand where commercial data added value and where they only added noise.

Analyses revealed that when compared to 2010 Census data, commercial data were useful for addresses and corroborating information in other files but had redundant information and little person coverage increase over Federal data. Addresses and person data were validated specifically. Person coverage matched what the Census found in Housing and Urban Development (HUD) and IRS databases. Some states (e.g., West Virginia and Wyoming) even demonstrated a 15 percent increase in addresses observed. On age and gender, SSA data were particularly accurate. On race, Census data were more accurate than any other source. Unfortunately, because of Title 13, this data cannot be shared with the HRS.

The Census Bureau did not use data compiling agencies such as MSG. Ultimately, it may have engaged too many vendors and, thereby, introduced more noise than signal from the mass volume of data. Supposedly, it received nine universe files. Consolidating nine files presented logistical and statistical challenges during the de-duplication process, particularly when considering that vendors stated that the files had already been de-duplicated.

From the Census Bureau perspective, the objective of its survey is to obtain data on all persons and addresses in all states. Combining Federal administrative data allows this. Name, address, place of birth, gender, and SSN are attached to the SSA Numident file. The quality of data for each data element varied across vendor and even over time (year over year and quarter to quarter) within vendors. Validation success seemed most highly-correlated with having full

identifiers: name, address, and date of birth (MM/YY, or MM/DD/YY). To resolve data conflicts, the Census Bureau aligned on using date stamps.

Records with protected identification keys (PIKs) were compared to MAF IDs. Overall, the study found about 91.2 percent of the 2010 Census over age 50 population was found in commercial data. This was higher than the overall (all ages) match of 88.6 percent. Commercial data covered similar percentages across older age groups, with over 90 percent for age groups 50 to 64, 65 to 74, and 75+.

From the HRS perspective, researchers need data for specific states and persons over age 50. A 2010 Census Match Study metadata analysis compared 2010 Census and commercial records for the population over age 50 by county. The data looked promising. Commercial data listed addresses where individuals were paying utilities or spending the most.

The Census Bureau also integrated other data: ACS tract level information, Medicare enrollment data, IRS 1040 data, phone numbers, and email. The ACS data are potentially problematic because they provide a 60-month average sample. It is not clear if the Census is doing a logical edit for every case. Medicare enrollment data keep older data strong. IRS 1040s describe wage earners: where they live and primary and secondary earners. Both location and filing status could be good data for HRS sampling. When the Census Bureau compared ACS to IRS data to define the distribution of income and wealth, it found the commercial data to be of low quality. Finally, the Census Bureau acquired phone and email data from vendors. Commercial data might have better data on phone numbers than Census data according to one analysis.

Opportunities and Tradeoffs for the Health and Retirement Study

Discussion led by John Thompson, NORC, and Roger Tourangeau, Westat

Citro opened the final discussion session with a brief recap of the day's presentations. First, sampling for rare populations is a concern for the HRS. There are various options for making screening more effective—for example, asking questions in a way that encourages accurate reporting. In screening surveys, researchers face a problem when respondents figure out what researchers are after (e.g., eligibility requirements) and inaccurately report to avoid response burden. Groves talked about challenges and opportunities to treating the MAF as a public utility. More cost-effective options for sampling included ABS, disproportionate share sampling, using a host survey, and purchasing unused screener samples from other surveys. She then posed the following question to the group: If the group had to rank order cost-effective methods that did not sacrifice methodological integrity, which would be the lowest hanging fruit? Which sampling methods should the research community prioritize and further develop?

Regarding commercial data, Fahimi was able to provide more insights. Depending on the data element of interest, some sources are better than others. Additionally, the more data elements that are included, the more the data are degraded in terms of quality. MSG relies on the highest-quality database for appending critical data items. If HRS researchers wanted to add age, without appending race, MSG data would be quite good.

Weir described the need for a long-term legislative agenda that will improve researchers' resources. Cost-effective sampling is not just about the HRS 2016 cohort. There is a longer-term problem that should be solved for researchers who need nationally representative samples. There is a certain obligation to conduct screening in a cost-effective manner. He asked how Title 13 could be revised to lift constraints on the MAF? Could researchers use IRS lists based on W2 data? These lists must have some demographic data to be useful.

Participants agreed that the HRS should make the best possible use of commercial data. The NIA has set aside some funds to do evaluations of sampling methods for the 2016 HRS screen. Some funds could be put toward evaluating one question that would help the HRS decide which vendor to enlist for address lists and which questions to ask.
Discussion continued around adoption of the NLSY cohort. The NLSY is preparing to add a younger cohort that better represents the ethnic composition of the nation today. It is less interested in following the older population. The Bureau of Labor Statistics (BLS) is planning to terminate all nonessential surveys and nonessential components of required non-discretionary surveys; the aging-in of the NLSY cohort to the HRS could be one way of keeping it going. If linked to the NLSY data, this comprehensive panel would offer a wealth of data for researchers to study. For instance, early cognitive data would be enormously valuable for understanding the evolution of cognition and onset of dementia. One issue with this panel, however, is the lack of recent immigrants. Another issue is that incorporation of the NLSY cohort is not guaranteed and therefore the HRS must pursue other options as well.

The HRS might consider collaborating with another survey screen. On the upside, it would be advantageous to pool resources to screen a larger national sample to obtain a household roster. Demand from RTI International, NORC, Michigan's Survey Research Center, and the NCHS (e.g., for NHIS, MEPS) could support such a broad survey screen. On the downside, researchers would have to agree on the screen and its timing. The cooperative agreement among survey research organizations would likely be tenuous. The data would only be useful for a limited time since people move.

The HRS should investigate how it could extend the life of a database by tracing and tracking households with eligibles or soon-to-be eligibles. Weir indicated that in past years, the HRS did not attempt to contact respondents between interviews but was considering this communication approach going forward on an experimental basis. Thompson advised that NORC had a policy of trying to keep respondents "warm" by sending periodic letters or emails. The NLSY no longer gets SSNs, which would aid in tracking respondents. Iams noted that the SSA gets its addresses from the IRS. The address list from the IRS may be the best source, which is derived from 1040 tax filings or W-2 data.

Weir also suggested that the HRS could conduct screening every 10 years to identify 46 to 56 year olds. The next cohort in 2016 would enroll 51 to 56 year olds. Then each year, individuals who turned 51 would be pulled into the survey from the pool of 46 to 50 year olds identified in the screen.

Kreuter emphasized that the HRS should ensure that fieldworkers know that they will be evaluated on both eligibility rates and response rates. Studying how interviewer knowledge

affects recruitment rates may be an interesting empirical question to pursue as well. The Census Bureau is building a database mapping response rates to the MAF in order to facilitate 2020 research and testing.

The HRS should employ multimode recruitment. Exhausting cheaper modes of contact (i.e., mail and telephone) prior to in-person contact would permit some savings. Focusing some effort on getting more telephone data could be worthwhile. The Census Bureau is investigating tranches of numbers that would be sent a text message with a code. It is further attempting to layer in additional data and constructing a more comprehensive database. Valliant noted that clustering does not work as well as expected. Segment stratification, particularly if one element is race, could still aid in blocking.

APPENDIX 1 – Meeting Agenda

Expert Meeting for the National Institute on Aging on Cost-Effective Household Screening Methods for the Health and Retirement Study

Committee on National Statistics
National Academy of Sciences
500 5th Street, NW, Washington, DC 20001
Keck Building, Room 101
November 19, 2012

The goal of this expert meeting is to discuss approaches to recruiting participants into the National Institute on Aging (NIA) Health and Retirement Study (HRS). Traditional methods of selecting primary and secondary sampling areas of neighborhoods in which to screen households to identify people in a desired age range are expensive. The expert meeting will consider alternative screening methods that could potentially be less expensive, but that could also assure a high-quality representative sample.

9:00–9:10 **Introductions and Welcome on Behalf of the National Academies**
Connie Citro, CNSTAT

9:10–12:05 **Screening Challenges Faced by the Health and Retirement Study**
Chair: John Thompson, NORC

- **9:10-9:45** **Background on the Health and Retirement Study and Rationale for the Meeting**
  John Phillips, National Institute on Aging
- **9:45–10:20** **Sampling Rare Populations**
  Graham Kalton, Westat
- **10:20–10:30** **Coffee Break**
- **10:30–11:30** **Sampling Challenges and Opportunities Specific to the U.S. Federal Statistical System**
  Robert Groves, Georgetown University
- **11:30–12:05** **Underreporting of Populations in Screening Surveys**
  Frauke Kreuter, University of Maryland

12:05–1:15 **Working Lunch to Continue Morning Discussions**
(NAS Cafeteria)

1:15–3:25 **Methodological Advances and Opportunities that Could Benefit the Health and Retirement Study**
Chair: Roger Tourangeau, Westat

1:15–2:00 **Ongoing Research on Alternative Screening Methods for the Health and Retirement Study**
Richard Valliant, University of Maryland

2:00–2:45 **Advances in List Based Sampling Methods**
Mansour Fahimi, Marketing Systems Group

2:45–3:25 **State-Based Address Lists**
Amy O'Hara, Census Bureau

3:25–3:35 **Coffee Break**

3:35–5:00 **Opportunities and Tradeoffs for the Health and Retirement Study**
Discussion led by John Thompson, NORC and Roger Tourangeau, Westat

5:00 **Adjourn**

APPENDIX 2 – List of Participants

Dallas Anderson, National Institute on Aging (NIA) Division of Neuroscience
Irena Dushi, Social Security Administration
Mansour Fahimi, Marketing Systems Group (MSG)
Robert Groves, Georgetown University
John Haaga, NIA Division of Behavioral and Social Research
Howard Iams, Social Security Administration
Graham Kalton, Westat
Frauke Kreuter, University of Maryland
Amy Mistretta, NIA Division of Behavioral and Social Research
Amy O'Hara, Census Bureau
John Phillips, NIA Division of Behavioral and Social Research
Richard Suzman, NIA Division of Behavioral and Social Research
John Thompson, National Opinion Research Center (NORC) at the University of Chicago
Roger Tourangeau, Westat
Richard Valliant, University of Maryland and University of Michigan
David Weir, University of Michigan

Committee on National Statistics of the National Academies Staff
Constance F. Citro, Director
Carol House, Senior Program Officer
Krisztina Marton, Senior Program Officer
Jacqui Sovde, Program Associate
Gooloo Wunderlich, Senior Program Officer