

B. Statistical Methods

1. RESPONDENT UNIVERSE AND SAMPLING METHODS

The target population for the FSS POMS is all civilian, non-institutionalized and over 18 years of age residents (citizens and non-citizens) of the United States.

As mentioned in Part A, core questions will be used to explore relationships among the concepts, develop a time series and measure any “shocks” to the system. By having a continual data collection, we will be able to look for changes in public perception after any of these types of events occur or look for underlying causes when we see a change in the time series. Based on past research, we know that attitudes towards administrative records usage and the census differ based on demographics (for recent examples see, Miller and Walejko, 2010, and Singer, Bates, and Van Hoewyk, 2011). We suspect that attitudes towards the federal statistical system may similarly vary by demographic. If shocks occur during this data collection, we need to be able to measure differences among demographic groups, particularly those demographics which have shown differences in past studies. To determine the necessary sample size, we examined the power necessary to detect a difference, should one exist, between demographic groups looking at rolled up weeks of data collection before and after a given event, or shock. Table 1 shows an example of a given daily or weekly demographic profile of the Gallup Daily Tracking survey.

	Daily (n=200)	Weekly (n=1400)	Adult Population Estimates (from 2011 Current Population Survey)
Gender			
Female	48.46%	49.15%	51.7%
Male	51.54%	50.85%	48.3%
Age			
18 - 34	17.72%	19.03%	30.5%
35 - 44	12.08%	11.94%	17.2%
45 - 54	19.34%	17.21%	19.1%
55 - 64	21.35%	21.50%	16.1%
65 +	29.51%	30.31%	17.1%
Education			
Less than HS	5.36%	5.80%	12.9%
HS Grad	18.87%	19.89%	30.2%
Some College	34.76%	33.85%	28.7%
College Grad	41.01%	40.46%	28.2%
Race			
White	84.71%	84.41%	82.4%
Black	9.43%	9.93%	12.1%
Asian	2.09%	2.19%	5.1%
American Indian	0.91%	0.62%	1.4%
Hispanic Origin			
Hispanic	7.55%	7.65%	13.6%
Non Hispanic	92.45%	92.35%	86.4%

Table 1. Example Daily and Weekly Profile of Gallup Daily Tracking Survey compared to Population Estimates from the 2011 Current Population Survey from the U.S. Census Bureau.

With 200 cases per night, analyses of change in public opinion for the entire sample are possible for shifts greater than plus or minus 10 percentage points at the 95 percent confidence level.

Routine analysis, however, will employ pooled weekly samples (n=1400). For week-to-week comparisons, pooled samples provide sufficient power to discern differences greater than 3 percent at the 95 percent confidence level. The sensitivity of measurement of change for subsamples, e.g. demographic groups, depends on the size of the subsample. In general, we do not plan to do subsample comparisons on a daily basis because the number of cases will be too small to detect significant changes. Weekly changes in opinion for different subgroups are

possible to detect in some cases, but changes for small subgroups (e.g. nonwhites, n approximately 170/week) will require more pooling of cases over days. Focusing on nonwhites as an example, pooling cases four weeks before and after a significant event, we could detect a 5 percent difference in attitudes. In sum, the design permits detection of large change in the full sample on a daily basis and smaller change for subgroups on a weekly or monthly basis.

Survey Design

The survey methods for the Gallup Daily tracking rely on live (not automated) interviews, dual-frame sampling (which includes listed landline frame combined with an RDD cell phone frame¹), and a random selection method for choosing respondents within the household. Additionally, the survey includes Spanish-language interviews for respondents who speak only Spanish, which includes sample coverage in Alaska and Hawaii, and relies on a three-call design² to reach respondents not contacted on the initial attempt. Nightly quotas exist to ensure that the un-weighted samples are proportionate by region. The data are weighted daily to compensate for

¹ The design currently used by Gallup for the G1K survey is a dual-frame design consisting of (i) a sample of listed landline numbers and (ii) a sample of cell phone numbers drawn from the telephone exchanges (dedicated exchanges) that are set aside for cellular providers. The landline sampling frame used for sampling of landline numbers therefore excludes unlisted landlines. Thomas M. Guterbock et.al. (Social Science Research 40 (Issue 3, May 2011) 860-872: "Who needs RDD? Combining directory listings with cell phone exchanges for an alternative telephone sampling frame") examines the feasibility of combining the EWP (Electronic white pages) sample with cell phone RDD (Random Digit Dialing), eliminating the ordinary RDD component from the sampling frame. This sampling method based on listed landline plus cell phones fails to cover only one segment of the telephone population: unlisted landline households that have no cell phone. They analyzed data from the 2006 National Health Interview Study to estimate the size of this segment and its demographic profile. Trend data from the NHIS were used to assess how these biases were changing. They found that the proposed alternative "Listed + cell" sampling frame provided relatively small bias compared to "RDD + cell" and the portion of the telephone universe that was excluded in the "listed + cell" design was getting smaller all the time, therefore its bias relative to the "RDD + cell" design was decreasing over time. Overall, the "listed + cell" design turned out to be a useful alternative. Based on these findings and data resulting from Gallup's internal studies, the "listed + cell" design was considered optimal for G1K sampling to increase overall sampling efficiency and was implemented starting April of 2011.

² Sampled cases will be in the system from when they are released until 3 call attempts are made and all 3 calls may not be made on the same day. Typically, they will stay in the system for 3 days unless they are resolved on the first or second call.

disproportionalities in selection probabilities and non-response. The data are weighted to match targets from the U.S. Census Bureau by age, sex, region, gender, education, ethnicity, and race. With the inclusion of the cell phone-only households and the Spanish-language interviews, more than 90% of the U.S. adult population is represented in the sample. By comparison, typical landline-only methodologies represent less than 70% of the adult population. The Gallup Daily Tracker response rate has averaged 11%, using an established formula for calculating response (CASRO).

Survey Sampling, Inc. provides listed landline sample and RDD cell phone sample (consisting of all exchanges set aside for cell phones) in non-overlapping frames³. The sample is stratified proportionately by census region. While the sample of 400 cell phone users is done without predetermined regional targets, the 600 landline surveys done by the Gallup Daily Tracking survey is stratified into four regions:

- East—130 completes;
- South—198 completes;
- Midwest—132; and
- West—140.

Through the Gallup Daily tracking survey, the U.S. Census Bureau will have access to obtain on average 200 nightly responses. Gallup will employ a randomized algorithm to ensure that the appropriate proportion (i.e., 20 percent) of interviews are selected nightly and are intermixed with the nightly Gallup target of 1,000 interviews. The nightly quota will reach the desired

³ The cell phone frame is for everybody (cell phone only or not). Any respondent sampled from this frame will be interviewed as long as he/she is otherwise eligible for the study. As noted before, the sampling frame consists of (i) listed landline frame and (ii) cell phone frame.

number on average. Thus, the sample size is 200 surveys of American adults every day for 350 days annually.

Gallup uses listed landline sample and RDD cell as their sample selection methodology. When calling a landline telephone, random selection is used to choose respondents within a household based on the most recent birthday. Cell phones are treated as *personal devices*, where the individual answering is considered to be the respondent, though Gallup does not currently ask any questions to determine whether or not this assumption is valid.. The *mode* by which Gallup collects data is to complete all surveys using an outbound phone mode.

The length of the survey is estimated at 10 minutes to complete the questions asked on behalf of the Census Bureau. Note that Gallup only uses completed surveys.

2. PROCEDURES FOR THE COLLECTION OF INFORMATION

Regarding data weighting, Gallup weights data daily to compensate for disproportionalities in selection probabilities. The data are further post-stratified using an iterative proportional fitting (i.e. raking) algorithm to account for nonrandom nonresponse by age, sex, region, gender, education, ethnicity, and race. The data are weighted to match targets from the U.S. Census Bureau by age, sex, region, gender, education, ethnicity, and race. With the inclusion of the cell-phone only households and the Spanish language interviews, over 90% of the U.S. adult population is represented in the sample⁴. By comparison, typical landline-only methodologies represent less than 70% of the adult population.

⁴ The sampling frame includes (i) listed landline frame and (ii) cell phone frame. The only segment that is not covered by this frame is the group of unlisted landlines without access to cell phones. The landline only accounts for roughly 9.2% and the segment consisting of unlisted landline numbers only (with no access to cell phones) is estimated to be around 3%.

Further, Gallup uses the latest available estimates from the National Health Interview Survey conducted by the National Center for Health Statistics to determine the target proportions by telephone status. (Gallup uses its prior experience with list-assisted landline surveys to determine listed/unlisted proportions.). Next, Gallup computes post-stratification weights based on targets from the Current Population Survey that the Census Bureau conducts for the Bureau of Labor Statistics. Next, Gallup uses an iterative proportional fitting (i.e., raking) algorithm to ensure the Gallup Daily tracking data match national targets for region by gender by age, age by education, race by gender, and ethnicity by gender. Finally, the company trims the weights to reduce variance so that the maximum range of the weights is no greater than 12 to 1. The survey documentation provides the weighting methodology (See Attachment D).

3. METHODS TO MAXIMIZE RESPONSE RATES AND DEAL WITH NONRESPONSE

Although the Gallup Daily Tracking Survey is portrayed as being nationally representative, it does not meet Census Bureau quality standards for dissemination and is not intended for use as precise national estimates or distribution as a Census Bureau data product. The Census Bureau and the Federal Statistical System will use the results from this survey to monitor awareness and attitudes, as an indicator of the impact of potential negative events, and as an indicator of potential changes in communication campaigns. The study that surrounded the 2010 Census demonstrates how these data can be useful for these types of decisions (Miller and Walejko, 2010). Miller and Walejko (2010) also demonstrate the usefulness of examining differences in public opinion among different demographics, such as race and age. Like the 2010 Census study, data from this research will be included in research reports with the understanding that the data were produced for strategic and tactical decision making and not for official estimates.

Research results may be prepared for presentation at professional meetings or in publications in professional journals to promote discussion among the larger survey and statistical community, encourage further research and refinement. Again, all presentations or publications will provide clear descriptions of the methodology and its limitations.

Gallup has a 51% contact rate, an 11% response rate (using the CASRO definition), 91% completion rate and a 22% non-interview/refusal rate. Gallup employs a three-call design to reach a respondent not contacted on the initial attempt and has the capacity to complete four callbacks.

Gallup calls individuals between the hours of:

- Monday to Thursday 4 p.m.–11 p.m. CST;
- Friday 3 p.m.–9 p.m. CST;
- Saturday 10 a.m.–3 p.m. CST; and
- Sunday 1 p.m.–6 p.m. CST.

Gallup generally completes the Daily tracking surveys between 9:30 a.m. and 10:30 p.m. CST. The call design ensures that each call after the initial one takes place at a different time in the afternoon or evening to maximize the likelihood of contact by dividing the interviewing period into 3 “buckets.” The three buckets are: early evening, late evening and weekend call times. For this study, each phone number will get 3 calls and efforts will be made to make those calls at various times (the three bucket times) of the day to the extent possible. Some numbers may get more than 3 calls. For example, if a call back request is made for a specific day and time on the

third call, that number may be called again to complete the interview. If a “busy” signal is detected, that number may be called after some time on the same day.

About Gallup⁵

Gallup trains its interviewers according to rigorous practices that have proven successful over the past 70 years of the company’s survey research work. In addition to full scope training on how to conduct the interviews before they begin their work, continuous training also takes place. This training combined with these interviewers’ experience and their above-average tenure allows Gallup to be uniquely successful at gaining trust and participation among the population.

Compared with industry averages, Gallup spends more time carefully recruiting, selecting, and training the best personnel for telephone interviewing assignments. In other words, the typical Gallup interviewer is better educated, better trained, more experienced, and more productive, on average, than interviewers in the survey industry as a whole. Because of this distinction, Gallup maintains one of the highest retention rates of interviewers in the industry, at more 2.2 years, on average, underscoring Gallup’s success in recruiting the right personnel for one of the most challenging jobs in survey work.

Gallup has been recognized for the application of this strengths-based approach in the selection and management of our entire staff of telephone interviewers. Our executive and consumer interviewers excel at the art of establishing a relationship instantly to put respondents at ease in relating their perceptions, feelings, habits, and attitudes. This is extremely important when interviewing respondents about their personal attitudes and activities.

⁵ The information in this section was provided directly by Gallup about their interviewing practices.

As indicated above, a psychological testing procedure developed and validated by Gallup is used in hiring to be certain that Gallup interviewers have the personality characteristics necessary to be successful as telephone interviewers. Gallup interviewers also must pass a screening interview that includes a reading test to evaluate voice quality, reading ability, and comprehension of questionnaire instructions.

The initial interviewer training consists of six hours of classroom instruction conducted by a training director over a two-day period. In addition, the training director works one-on-one as an additional supervisor with new interviewers individualizing on-the-job training and evaluation for the first six weeks. During this period, the training director reviews completed interviews, which have been tape-recorded with the interviewer, so that instruction is concrete and personal.

Nonresponse Bias Study

Through its Gallup Daily tracking survey, the company will sample up to 10,000 non-interviews to re-contact over the course of a year. Non-response interviews do not include hard refusals, which by company standards Gallup does not re-contact. This will include landline and cell phone sample. Gallup will call each non-interview record up to 15 times in an attempt to make contact and gain cooperation. This means that Gallup could call respondents up to 18 times including the initial three calls completed. Gallup will attempt to complete an abbreviated questionnaire (i.e., standard demographics and the approximately 19 core Census Bureau survey questions) with this sample. This study will be conducted quarterly between the periods of February 2012 and January 2013 (a total of 4 quarters with a sample up to 2500 in each quarter). Results from these will be delivered quarterly to the Census Bureau.

Gallup will provide the U.S. Census Bureau only the data collected through the U.S. Census Bureau questions and the demographic questions. Note that the purpose of this study includes a comparative analysis of the response patterns of non-response participants and typical participants. The goal of the analysis plan is to compare the respondents and the non-respondents and, based on those findings, examine the nature of non-response pattern and potential of non-response bias. As mentioned in the document, Gallup will sample up to 10,000 non-interviews continuously to re-contact over the course of a year. Non-response interviews will not include hard refusals but will include both landline and cell phone sample. Gallup will call each non-interview record up to 15 times in an attempt to make contact and gain cooperation. Gallup will attempt to complete an abbreviated questionnaire (i.e., standard demographics and the approximately 20 or so core Census Bureau survey questions) with this sample.

The distribution of the key demographic variables among respondents and non-respondents will be compared to examine if certain demographic subgroups were more or less likely to participate in the survey. Similar comparison will also be made for each of the core Census Bureau questions to identify which, if any, of these questions were subject to potential non-response bias. It is expected that the non-response follow-up study based on a sample size of up to 10,000 non-respondents over the course of a year will yield between 300 and 1500 interviews. The response rate for the non-response follow-up study is expected to be between 5 - 10%.

4. TEST OF PROCEDURES OR METHODS

The Organization for Economic Co-operation and Development (OECD) electronic working group on measuring trust in official statistics developed a survey for measuring trust in official

statistics that was cognitively tested in six of the member countries including the United States (Brackfield 2011). The goal of this development was to produce a model survey questionnaire that could be made available internationally to be used comparably in different countries.

Unfortunately, a 2010 National Center for Health Statistics (NCHS) cognitive study revealed that these questions are inadequately understood by U.S. respondents (Willson et.al. 2010) and therefore would be unable to sufficiently measure the trust in the FSS in the United States. As such, the FSS Working Group sought to build upon the theoretical constructs and previous research on this subject (Felligi 2004; OECD Working Group 2011; Willson et.al. 2010) in designing and administering a version of this poll that might adequately measure U.S. public opinion of the FSS. Those questions that were shown to work well in the NCHS cognitive test were included in the survey for cognitive testing.

The Federal Statistical System (FSS) Team focused on definitions of trust in statistical products and trust in statistical institutions that are derived from work by Ivan Fellegi (1996, 2004). In addition to considering questions developed and tested by the OECD working group, the FSS working group considered questions used by the Office of National Statistics and the National Centre for Social Research in the United Kingdom and by the Eurobarometer. Based on the U.S. cognitive laboratory results from NCHS work on the OECD survey, we started from the premise that we needed to measure awareness of statistics and statistical institutions first, and then assess level of knowledge/data use, before proceeding to questions addressing trust. We consulted additional previous research that examined the U.S. public's knowledge of statistics (Curtin, 2007) and sought to create a questionnaire that would be comprehensible by the general population. More detail on questionnaire development is available in Childs, et al. (forthcoming).

The second overarching goal of this public opinion effort was to gauge public opinion towards the use of administrative records for statistical purposes. For this part of the questionnaire development, we reviewed previous questions on the topic, mostly from the perspective of the Census Bureau (Miller and Walejko, 2010; Singer, Bates and Van Hoewyk, 2011; and Conrey, F.R., ZuWallack, R., and Locke, R., 2011). We also considered work conducted internationally. A 2009 study conducted by the ONS revealed that the UK general public is varied in their knowledge about government agencies and their current levels of data sharing. Over fifty percent of respondents were aware that no single government central data base currently exists, but that there are separate databases maintained by individual departments, though this varied by education, age and region. Overall the response received was supportive (approx. two thirds in favor) of data sharing and the creation of a single central population database of UK residents. By including similar questions about knowledge and evaluations of data sharing, the FSS may be able to take measures to increase awareness and/or alter current data sharing practices, which would enable the government to save costs and improve data quality.

Cognitive Testing

The Census Bureau, NCHS, and IRS conducted 42 cognitive interviews of a purposive sample that was diverse in terms of age, race, gender, education and trust in the government (see Willson, forthcoming for more information). The driving factor shaping the question-response process in these questions was respondents' lack of understanding and knowledge of the FSS in particular and statistical information in general. This is consistent with findings from an OECD international effort aimed at measuring attitudes about the general population's trust in statistics produced by national governments which found that respondents have very little knowledge of

official statistics. As a result, the survey questions fall victim to a phenomenon common among many attitude questions; that is, there is no static underlying evaluation to capture. Instead, responses are created on the spot and are often inconsistent across questions aiming to measure similar concepts. Four themes that illustrate such instability and absence of predefined opinions about federal statistics are outlined below.

1. Shifting interpretations: There were many instances of respondents giving contradictory answers to similar survey questions or providing a narrative during probing that did not match the way they answered the survey question. There were many examples of this pattern in the data. It is apparent that people's answers easily fluctuate, depending on what they are prompted to think about and what considerations they sample before answering.

2. Not thinking about statistics: Many people do not have predefined opinions about federal statistical data because they have little awareness of this type of information to begin with. Even those with some level of awareness do not possess sophisticated knowledge of statistical information and have not thought extensively about the topic, especially in relation to the Federal government. Some respondents (those with no knowledge of federal statistics) do not interpret the questions as intended, and, as a result, the statistics produced by these items will not reflect the desired construct.

3. Confusion over what the question was asking: A related point is that respondents with very little knowledge of federal statistics sometimes had difficulty understanding what a question was asking altogether.⁶ If this confusion was great enough, they could not determine what beliefs to sample and, therefore, could not answer the question at all. Many times, respondents would

⁶ A lack of awareness of federal statistics seemed to be correlated with the educational level of our respondents, but this cannot be statistically confirmed by this qualitative study.

draw upon their own experience in relation to the topic (not statistics on the topic). For example, one question states, “Information collected to create federal statistics is sometimes used by the police and the FBI to keep track of people who break the law.” Many respondents were not thinking of statistics as much as they were thinking about the government accessing people’s personal files or police records. They cited examples such as terrorists lists, sexual predator lists, lists of traffic ticket recipients, travel records and personal files.

4. Interpretations limited to examples given: A final indication that respondents cannot think generally about federal statistics is that many respondents limited their interpretation of the questions to the examples given in the first question, especially if they were already familiar with the agency. (The Census Bureau was, by far, the most recognized agency, but others were mentioned as well.) Even when subsequent questions asked about federal statistics more generally, some respondents thought specifically of the agency they knew, such as Census.

Proposed Solutions: Define the Context, Be Specific, Keep it Simple

Because most people in the general population have little or no knowledge of the FSS and have given little thought to statistical information in general, these attitude questions, like many attitude question, are likely to produce unstable estimates. The challenge, then, for question design is to craft questions that elicit consistent interpretations among respondents. We suggest three question design strategies for improving construct validity.

1. Define concepts up front: Because many people do not have a good deal of knowledge about federal statistics, it’s important that question convey this topic right away and consistently. The first tested question does a good job of setting the context for the rest of the questions to the

extent that it uses specific federal statistics. This turned out to be the most important function of question 1, and the question could be strengthened in this regard by eliminating non-federal statistics. Rather than a test of knowledge (which we already know is limited), it should serve as a “primer” to define what it is we mean by federal statistics and federal agencies.

2. *Be specific:* Many respondents in our sample were thinking of specific federal agencies when answering the questions. The most common agency cited was the Census Bureau, but others were mentioned too, such as the Bureau of Labor Statistics. Similarly, respondents tended to think specifically about unemployment statistics or the population count even when the question was asking them to think generally. The downside to thinking of specific examples is that the question is measuring opinions on those examples only. However, we contend that this is preferable to respondents not thinking about federal agencies or statistics at all – which tended to happen when they were presented with broad topics such as “federal statistics.”

3. *Keep it simple:* Each question should be as straightforward as possible and avoid complex concepts that require higher-level thought and analysis on the part of respondents. Many respondents are not aware of or have given much thought to federal statistics. Questions should not mix concepts or present complicated scenarios. Nor should they contain broad concepts that invite multiple understandings.

These findings, as well as question-specific findings were incorporated into the questionnaire that was pretested for three weeks. Results from that pretest will be fielded for this survey. Attached is the set of questions that went the field pretest described below (Attachment C).

Field Pretest

The field pretest consisted of three weeks of pretesting. Each week corresponds to a different phase. This pretest was covered under the Generic Clearance for Pretesting, but is described here for comprehensiveness.

During Phase One, 30 items were piloted and evaluated using a variety of methods: 1) Item response distributions were assessed to determine the value of various response categories; 2) The factor structure of the scaled items was explored using Exploratory Factor Analysis (EFA), which allowed us to determine if items loaded under the factor they were designed to measure, if items load under multiple factors (cross-load), and if the hypothesized factors were identified (meaning at least three items load under them); and 3) Random probing was used to increase our understanding of how each question was being interpreted. Using results from the above analysis, the following changes were made to the questionnaire for Phase Two:

- Awareness items – We dropped the first three questions that asked if a person had heard of the statistic because we found that most respondents had heard of these very common statistics, and when they responded that they had not, they reported in random probes that they did not know the figure, but they had heard of the statistic. Instead, we asked knowledge of who measures each item and we added three additional statistics to increase our variance.
- Agree/Disagree items –We added a midpoint to the agree/disagree scale (neither agree nor disagree). This was based on the probe data suggesting little differentiation between somewhat agree and somewhat disagree. Also literature that suggests that a midpoint is advisable.

During Phase Two, 30 items were piloted and evaluated in terms of item and model fit (also shown in Attachment C). Using a theoretical model in Confirmatory Factor Analysis (CFA) we assessed measurement error, model invariance, and reviewed recommendations for improving model fit by either eliminating items or respecifying the factor structure. We reexamined random probes to look at the impact of changes made after Phase One. Based on these analyses, items were kept using the following criteria:

- Good understanding demonstrated with cognitive testing and random probes,
- Good fit to conceptual factor model based on a conceptual understanding of how the item should load (iterative process),
- Allows international comparison, and/or
- Only item to measure a concept.

During Phase Three, 25 questions were fielded that met the above criteria (also shown in Attachment C). Random probing was used for items that were changed after Phase Two and for items that had demonstrated some problems in Phase Two. After Phase Three, only one question was dropped, which was found not to function ideally. In its place, an additional “confidence in institutions” item was added on universities. Attachment A shows the final questionnaire.

5. INDIVIDUALS CONSULTED ON STATISTICAL ASPECTS AND INDIVIDUALS COLLECTING AND/OR ANALYZING DATA

The external statistical consultant and contactor is:

Rajesh Srinivasan, Gallup (609.279.2554).

Within the Federal Government, consultants include the Statistical and Science Policy Office, Office of Management and Budget; National Agricultural Statistics Service; the National Center of Health Statistics the Economic Research Service and the Statistical of Income Division, IRS. Staff from the Center for Survey Measurement in the Census Bureau and from the Bureau of Labor Statistics will analyze the information collected.

References

Brackfield, D. (2011.) “OECD Work on Measuring Trust in Official Statistics”

Childs, J.H. (forthcoming). Federal Statistical System Public Opinion Survey Study Plan.

Conrey, F.R., ZuWallack, R., and Locke, R. (2011). Census Barriers, Attitudes and Motivators Survey II Final Report, 2010 Census Program for Evaluations and Experiments, October 31, 2011.

Curtin, R. (2007), “What US Consumers Know About Economic Conditions”, paper presented at the second OECD Workshop on “Measuring and Fostering the Progress of Societies, Istanbul, June 27.

Data, A. R., Yan, T., Evans, D., Pedlow, S., Spencer, B., and Bautista, R. (2011). 2010 Census Integrated Communications Program Evaluation (CICPE) Final Evaluation Report. August 3, 2011.

Fellegi, I. (1996). Characteristics of an effective statistical system. Canadian Public Administration, Vol 39, Issue 1, pg. 5-34.

Fellegi, I. (2004). “Maintaining the Credibility of Official Statistics” in *Statistical Journal of the United Nations* ECE (21) 191–198.

Fellegi, I. (2010.) “Report of the electronic working group on measuring trust in official statistics”, *OECD Meeting of the Committee on Statistics*, June 2010, Paris.

Miller, P.V., Walejko, G. K. (2010). Tracking Study of Attitudes and Intentions to Participate in the 2010 Census. Technical Report submitted to the U.S. Census Bureau for Contract YA1323-09-CQ-0032, Task Order 001, December 10, 2010.

OECD Working Group. (2011.) “Measuring Trust in Official Statistics—Cognitive Testing”

Willson, S., Ridolfo, H., and Maitland, A. (2010.) “Cognitive Interview Evaluation of Survey Questions Measuring Trust in Official Statistics”

Singer, E., Bates, N. & Van Hoewyk, J. (2011). Concerns about privacy, trust in government, and willingness to use administrative records to improve the decennial census. Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Phoenix Arizona.