

B. Statistical Methods

1. RESPONDENT UNIVERSE AND SAMPLING METHODS

The target population for the FSS POS 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 weekly demographic profile of the Gallup Daily Tracking survey.

	Demographics	Weekly (n=850)
	Female	49.6%
	Male	50.4%
	18 - 34	16.3%
	35 - 44	11.2%
	45 - 54	15.5%
	55 - 64	20.7%
	65 +	36.3%
	Less than HS	4.2%
	HS Grad	27.0%
	Some College	23.2%
	College Grad	45.6%
	White	85.6%
	Black	7.9%
	Asian	1.8%
	American Indian	1.3%
	NHPI	0.4%
	More than one Race	3.6%
	Hispanic	5.5%
	Non Hispanic	94.5%

Table 1. Example Weekly Profile of Gallup Daily Tracking Survey.

With 850 cases per week, analyses of change in public opinion for the entire sample are possible for shifts across weeks that are greater than plus or minus 10 percentage points 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. 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 10 percent difference in attitudes. In sum, the design permits detection of large change in the full sample on a weekly basis and changes for subgroups on a monthly basis.

Survey Design

Survey Sampling Inc. provides Gallup with listed RDD list-assisted landline sample and RDD wireless phone sample (consisting of all exchanges set aside for wireless phones) in non-overlapping frames. The RDD list-assisted landline and RDD wireless phone samples are stratified by U.S. Census region and by time zone within region.

Gallup completes 500 wireless phone surveys and 500 landline surveys daily divided evenly between the two topical questionnaires. Gallup stratifies the samples into seven time zone within region geographies:

- East — 92 landline completes, 92 wireless phone completes
- Midwest region, Eastern time zone — 44 landline completes, 44 wireless phone completes
- Midwest region, Central time zone — 66 landline completes, 66 wireless phone completes
- South region, Eastern time zone — 102 landline completes, 102 wireless phone completes
- South region, Central time zone — 80 landline completes, 80 wireless phone completes
- West region, Mountain time zone — 30 landline completes, 30 wireless phone completes
- West region, Pacific time zone — 86 landline completes, 86 wireless phone completes

Note that geographies that fall outside the dominant time zones in a region are rolled into the next closest time zone within the same region for sampling, sample management and weighting purposes. For instance, the Midwest region Mountain time zone is rolled into Midwest Central,

South region Mountain time zone is rolled into South Central, and West Region (west of the Pacific) is rolled into West Pacific.

Gallup uses an RDD list-assisted landline sample and an RDD wireless phone sample. When calling a landline telephone, Gallup uses random selection to choose respondents within a household based on the most recent birthday. Gallup treats wireless phones as personal devices: The individual who answers the wireless phone is the respondent.

Sample Sizes: Gallup conducts 1,000 surveys with American adults, aged 18 and older, daily, 350 days annually. Five hundred respondents are asked the Wellbeing track survey, while the other 500 complete the Politics and Economy track survey. Census questions are going to appear on the Politics and Economy track survey. Certain variables, such as employment indicators and key demographics, are asked on both survey tracks.

2. PROCEDURES FOR THE COLLECTION OF INFORMATION

To compensate for unequal selection probabilities and nonresponse, Gallup weights the data daily. Gallup further post-stratifies the data using an iterative proportional fitting (ie: raking) algorithm to account for nonresponse by phone status, age, sex, region, education, population density, ethnicity, and race. Gallup calculates daily weights for each survey track separately and for the combined data. Gallup computes state weights twice per year and congressional district and MSA weights once per year.

The following procedures are used:

1. Calculate selection probability and nonresponse weights to compensate for disproportionalities in probabilities of selection and response rate by sample frame. These are calculated separately by time zone within region within the RDD landline phone sample and then within the wireless phone sample.
2. Calculate selection probability weights to compensate for unequal probabilities of selection for respondents reached via a landline phone. Because Gallup only interviews one adult per landline household, these weights are based on the number of adults in the household.
3. Calculate selection probability weights using the lambda compositing method for dual phone users. This makes necessary adjustments in probabilities of selection for respondents who have both a landline and a wireless phone, and thus could be in both the landline and wireless phone sample frames, versus respondents with the possibility of being in only one sample frame. Gallup calculates the dual-user weights to account for the proportion of dual users from the landline versus the wireless phone sample frames. Each respondent's dual user status is based on whether their household has a landline phone and they personally have a wireless phone.
4. Use an iterative proportional fitting (i.e., raking) algorithm to ensure the Daily Tracking data match national targets of telephone status, Census region by age, gender by age, education, race by Hispanic ethnicity, and population density quintile of self-reported county.
 - a. Gallup calculates post-stratification weights for telephone status using the latest available estimates from the National Health Interview Survey conducted by the National Center for Health Statistics to determine the individual-level target

proportions by household telephone status. While this is an individual level weight and individual level weighting targets are used, each respondent's telephone status for this weight is based on their household's telephone status. This is done in order to match the method used to define individual's telephone statuses by the NHIS.

- b. Gallup calculates demographic post-stratification weights based on targets from the Current Population Survey the U.S. Census Bureau conducts for the Bureau of Labor Statistics.
 - c. Gallup calculates population density weights based on targets from the Decennial census.
5. Trim the final weights to reduce variance.
 6. Calculate weights for each track separately and for the combined data

Through the Gallup Daily tracking survey, the U.S. Census Bureau will have access to obtain on average 121 nightly responses, aggregating to 850 per week. Gallup will employ a randomized algorithm to ensure that the appropriate proportion of interviews are selected nightly and are intermixed with the nightly Gallup target of 1,000 interviews. The nightly quota will reach the desired number on average. Thus, the sample size is 850 surveys of American adults every week annually.

The length of the survey is estimated at a maximum of 10 minutes to complete the questions asked on behalf of the Census Bureau. On most nights, the census questions will not exceed 5 minutes. Note that Gallup only uses completed surveys.

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.

Length of Survey: The Politics and Economy track interview currently takes participants an average of 8 minutes. As Gallup adds questions, interview times can vary but never exceed 18 minutes on average. The 18-minute limit, based on Gallup research, is set to restrict attrition during the course of the interview, as the longer the survey, the more likely a respondent will not complete the interview in its entirety. Gallup only uses completed surveys.

Contact rate: 39%

Response rate: 11% (AAPOR III Response Rate)

Completion rate: 93%

Non-interview/Refusal rate: 28%

Number of Callbacks: Gallup uses a three-call design to reach the sampled household and complete an interview with the randomly selected respondent. Gallup may make more than three attempts on a number under certain circumstances such as a scheduled callback.

Gallup is not planning to perform nonresponse bias analysis other than standard post-stratification that will be performed as part of the nightly weighting.

Gallup calls individuals between the hours of:

- Monday to Thursday 4 p.m.–11 p.m. CST
- Friday 4 p.m.–9:30 p.m. CST
- Saturday 10 a.m.–3 p.m. CST
- Sunday 1 p.m.–6 p.m. CST

Gallup generally concludes the daily surveys between 9:30 p.m. and 10:30 p.m. CST Monday to Friday. 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 three “buckets.” Monday to Friday,

Gallup calls individuals between 5 p.m. and 9 p.m. in respondents' time zone unless they request a callback. Saturdays, Gallup begins dialing at 10 a.m. in each time zone, and on Sundays, Gallup begins dialing at 11 a.m. in each time zone.

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.

4. TEST OF PROCEDURES OR METHODS

In the prior iteration of this study, we found that (1) data users are more likely to report trusting statistics than non-data users; (2) reported belief in the credibility of statistics predicts reported trust in federal statistics; (3) reported trust in statistics remained relatively stable over the two year data collection; and (4) the government shut-down caused by a deadlocked Congress coincided with the largest dip in reported trust in statistics (however, this may have been confounded by distrust of the roll out of the Affordable Care Act). These findings contribute to

ongoing conversations about how to improve response rates and efficiency in the FSS. Some of these findings are included in Attachment D.

Past testing using this panel has looked at opinions of the use of administrative records in lieu of a survey response, opinions on preferred survey response options and opinions on using your own device for work purposes.

Questions regarding administrative record use has shown that respondents tend to prefer being asked questions on surveys to the use of records; however, when the use of administrative records is framed to indicate to the respondent that the use of records can save the government money or provide a social good then respondents are more likely to favor using administrative records. Findings also seem to indicate that respondents prefer the use of government records to “public” or other third-party records. Some of these findings are demonstrated in Attachment E.

Exploration of context effects has shown that ratings of trust in federal statistics are affected by the preceding questionnaire content. The Gallup Daily Tracking Survey, on which the FSS POS items are collected, is divided into two topical paths or tracks. The first track, known as the Wellbeing Index (WB), asks respondents about topics related to health and well-being. The second track, known as the Political & Economic track (P&E), asks respondents for their opinions about various political and economic topics. When five core FSS POS items about trust in federal statistics were included concurrently on both survey tracks for six weeks, responses showed that respondents who answer the FSS POS items in the context of a survey about politics and the economy (1) were less likely to report being users of federal statistics; (2) were less likely to trust federal statistics ; (3) were less likely to agree that policy makers need federal

statistics; and (4) were less likely to agree that the FSS keeps data confidential compared to respondents answering in the context of questions about health and wellbeing.

Data from the first and second contracts with Gallup, and the prior OMB clearance, were collected on the WB track of the Gallup survey. In the current contract, the FSS POS is being collected on the P&E track of the Gallup survey. The concurrent data collection discussed above indicates what differences should be expected as a result of this change. There is no reason to predict that the relationship between items will change when data are collected on the P&E track, but the item-level responses will most likely change, reflecting overall slightly lower levels of trust. For this reason, future time series will not directly compare prior responses collected on the WB track with new responses collected on the P&E track.

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.