GALLUP®

June 4, 2013

GALLUP 2012 PRESIDENTIAL ELECTION POLLING REVIEW

TABLE OF CONTENTS

COPYRIGHT STANDARDS

This document contains proprietary research, copyrighted materials, and literary property of Gallup, Inc. It is for the guidance of your company only and is not to be copied, quoted, published, or divulged to others outside of your organization. Gallup®, The Gallup Poll®, and Selection Research, Inc.™ are trademarks of Gallup, Inc. All other trademarks are the property of their respective owners.

This document is of great value to both your organization and Gallup, Inc. Accordingly, international and domestic laws and penalties guaranteeing patent, copyright, trademark, and trade secret protection protect the ideas, concepts, and recommendations related within this document.

No changes may be made to this document without the express written permission of Gallup, Inc.

GALLUP 2012 PRESIDENTIAL ELECTION POLLING REVIEW

Gallup conducted a thorough review of its 2012 pre-election presidential polling to determine the factors that affected its 2012 results. Gallup Editor-in-Chief Frank Newport led the review process, along with Dr. Michael Traugott of the University of Michigan and a team of Gallup's most experienced statisticians, methodologists, and research analysts. Dr. Frauke Kreuter of the University of Maryland, Dr. James Wagner of the University of Michigan, and Dr. Christopher Wlezien of the University of Texas provided advice and consultation.

The extensive review process involved a significant amount of new research, including the fielding of experiments and simulations focused on three areas of pre-election polling -- survey and sample design, survey field management, and data handling. Gallup has identified specific factors within these three broad areas that it believes were core contributors to Gallup's 2012 presidential election polling issues. Gallup's election polling review and experimentation process is ongoing, and Gallup will continue to add to this body of research in the months ahead.

INTRODUCTION

In the 2012 U.S. presidential election, most pre-election polls underestimated Barack Obama's popular vote strength. President Obama won the popular vote by 3.85 percentage points over Mitt Romney, while the average of the major polls using a landline and cellphone methodology estimated President Obama would win by about one percentage point. There was a distribution of individual poll estimates around this average.

Gallup's final pre-election estimate of the 2012 popular vote, based on likely voters, was Romney 49% and Obama 48%, with a margin of error of ±2 percentage points for each candidate's estimate. Gallup's goal in presidential election environments is to reflect underlying voter opinions and sentiment accurately, and in its final estimate, to closely approximate the final popular vote outcome. To determine why its final estimate was different from the actual result, Gallup explored a variety of factors related to its pre-election polling.

Gallup has thus far concluded that four of these factors most likely contributed to the difference between Gallup's estimate and the final results.

Changes have been implemented or are currently being implemented, and in one instance, continuing research is being conducted, to address these four factors:

1. **Gallup's likely voter estimating.** Gallup's likely voter procedures in 2012 shifted the race four points in Romney's favor; one point more in Romney's favor than the average shift among other polls for which registered voter and likely voter information is available. Gallup's procedures are broadly similar to those of other survey organizations, most of which sort respondents into likely voters and non-likely voters based on self-reported turnout questions. Gallup's likely voter questions are more heavily weighted toward past voting behavior than other firms' questions, although reducing or removing past voting from the model would not alone have made Gallup's final estimate more similar to other firms' estimates. The more significant variable in Gallup's model that pushed the vote share more toward Romney was the "thought given to the election" variable -- whose removal, along with other changes in the way the likely voter figures were calculated, would have made Gallup's final estimate more similar to those of other firms.

   Gallup will continue to investigate its likely voter procedures and explore whether major revisions or even a replacement of the model is needed. Continuing changes to the election system and to the polling industry -- such as declining response rates, increases in early voting, more intensive campaign efforts to reach voters, and penetration of new technologies -- will likely necessitate continuing changes in estimating likely voters and election outcomes. Gallup will conduct a complete re-evaluation of its likely

voter procedures in the coming months. Gallup will work with the University of Michigan Program in Survey Methodology in the fall of 2013 to plan and conduct a series of in-depth experiments of likely voter estimating procedures, using the 2013 Virginia and New Jersey gubernatorial elections as real-world test cases.

2. **Gallup's regional controls on interviews.** Gallup attempts to ensure geographical representation of the U.S. population in its national samples by requiring a minimum number of completed interviews by census region, and through weighting by region. However, interviewing factors can influence the percentage of completed interviews obtained in geographic units within each region. As a result, completed interviews by time zone within a region can be disproportionate, even if a region is appropriately represented. At times during the pre-election polling period, this led Gallup to underrepresent the Eastern time zone within the Midwestern and Southern regions, and the Pacific time zone within the Western region. In addition, regional stratification was not used to release cellphone sample until further into the pre-election polling period, meaning the cellphone sample was completed too quickly and inconsistently across regions.

    Simulations indicate that underrepresentation of specific time zones within regions led, by varying degrees at differing time periods, to Gallup's underestimate of Obama's vote share during the Oct. 1-Nov. 4, 2012, presidential election tracking. Gallup is currently working with both internal and external experts to develop appropriate sample control mechanisms for use in its political and economic surveys, and will be implementing these changes over time.

3. **The way in which Gallup asked about and weighted race and ethnicity.** In its pre-election polling, Gallup collected race and ethnicity information from respondents, using a series of forced-choice yes/no questions for each of several race and ethnicity categories, which resulted in a disproportionate number of respondents reporting they were multiracial and American Indian/Alaska Native. Because Gallup weights its data according to race targets established by the U.S. Census Bureau's Current Population Survey (CPS), accurate representation of all ethnic and racial groups is highly important. In its pre-election polling, Gallup weighted its data by two racial or ethnic categories -- Hispanic and black.

    In early 2013, Gallup addressed these issues by changing the way it collects race and ethnicity data; respondents are now read a list of all racial and ethnic categories at once, and respondents can select up to five categories. This change in race/ethnicity measurement has been accompanied by modifications in weighting procedures, which now involve weighting to four ethnic and racial categories. The impact of these changes will be monitored over time. Gallup's presidential job approval ratings and the average of other national polls' ratings have come into closer alignment in the first months of 2013.

4. **The use of a listed landline instead of an RDD list-assisted landline sample frame.** Gallup's experimental research in the spring of 2013 showed that on an unweighted basis, the listed landline sample in use in 2012 consisted of older and more Republican respondents than the RDD list-assisted landline sample. The listed landline and cellphone sample frames also underrepresented dual cellphone and landline users whose landline was unlisted, which required significantly more weighting of respondents interviewed on cellphones. Additionally, those who were included in the listed landline sample were demographically different than respondents from the RDD list-assisted landline plus cellphone sample frames. These differences likely contributed to Gallup's less accurate vote estimate. Gallup has addressed this issue by making the transition back to an RDD list-assisted landline sample, which is now underway.

While these are the main factors Gallup has identified thus far, further research is being conducted in additional areas, including the random within-household selection procedures used to select respondents from sampled households. In-depth research on the appropriate call design and number of calls per sampled phone number has also been completed and is now being analyzed.

Additionally, given that nearly all of the major national media and independent polls underestimated Obama's final popular vote margin to some degree, there were factors operating in this campaign that affected the entire survey research industry in 2012. Election administration procedures are continually changing; political campaigns are becoming more sophisticated and highly targeted; and polling methods face continuing challenges in data collection and lower response rates. Because all pre-election pollsters face a number of similar issues in estimating the outcome of contemporary elections, Gallup is committed to making the results of this research effort widely available for extended analysis by others.

Gallup's ongoing measures of unemployment, economic confidence, and other measures closely match those put out by the government and by other organizations, and in the first months of 2013, Gallup's presidential job approval measure has been closely aligned with the average of other national polling organizations' measures. But the estimation of presidential election outcomes is a particular challenge, and trying to reconstruct factors that might have affected any firm's estimation in an election campaign is sometimes difficult because the exact environment surrounding the election cannot be re-created retrospectively. However, this review has been very informative in identifying specific factors that could have affected the 2012 presidential polling and in identifying hypotheses that can be tested prospectively with additional research. In addition, this review has provided extensive information about the polling process to both Gallup and the greater survey research community.

SUMMARY OF RESEARCH AREAS

A. SURVEY AND SAMPLE DESIGN FACTORS

1. TRACKING DESIGN

Issue: Gallup's election polling uses a tracking design that mandates a fixed number of completed interviews per night, overall and in specified groups. By comparison, a stand-alone design that is conducted over the course of several days can include a variable number of completed interviews per night. In the tracking design, each night's sample is self-contained, which means it is weighted on a nightly basis. The tracking design requires daily weights based on smaller sample sizes than is the case for weighting done on multinight samples.

Research conducted: To test the implications of this tracking design on Gallup's 2012 pre-election polling, Gallup conducted a simulation in which the base data for seven-day weeks were weighted as an overall seven-day aggregate. The aggregate estimates were then compared with the results when the same data were weighted as seven individual samples.

Findings: The results showed no significant differences in the resulting distributions of presidential vote choice.

Conclusion: The results of this simulation indicate that the weighting implications of a tracking design, by themselves, were not a factor that caused Gallup to underestimate Obama's vote share in the 2012 presidential election. The impact of other operational requirements of a daily tracking design will be discussed in following sections of this report.

2. RDD LIST-ASSISTED LANDLINE VS. LISTED LANDLINE SAMPLES

Issue: From Oct. 1-Nov. 4, Gallup used a dual-frame telephone sample design consisting of a 50% RDD cellphone sample and a 50% listed landline sample. Gallup began using a listed landline sample instead of an RDD list-assisted landline sample concurrent with an increase in the percentage of interviews conducted by cellphone in April 2011. At that time, research suggested that the use of a larger cellphone component and listed landline sample provided coverage of all households in the United States, with the exception of a small percentage of those with an unlisted landline and no cellphone, and those with no telephone at all.

Research conducted: After the 2012 presidential election, Gallup carefully re-evaluated the use of the listed landline sample. An experiment was conducted with two side-by-side samples; one survey used a 50% listed landline sample and 50% RDD cellphone sample, and the other survey used a 50% RDD list-assisted landline sample and 50% RDD cellphone sample.

Findings: There were minimal differences in the weighted data resulting from the two studies in terms of respondents' political attitudes and their demographic characteristics. However, on an unweighted basis, the RDD list-assisted landline sample was younger, had different phone status categorization, was more Democratic, and was more likely to approve of President Obama than the listed landline sample. Additionally, in the listed landline and RDD cellphone combined sample, the respondents with both an unlisted landline and a cellphone were underrepresented and demographically different from the unlisted dual users in the RDD list-assisted landline sample and RDD cellphone combined sample. Thus, the cellphone component of the listed landline and cellphone sample was assigned significantly higher weights when compared with the cellphone sample in the RDD list-assisted landline and RDD cellphone combined sample. The significantly higher weights attributed to the cellphone component in the listed landline and RDD cellphone sample resulted in a higher design effect than the design effect for the RDD list-assisted landline and RDD cellphone combined group.

Conclusion: Although the weighted differences in certain outcome variables in the spring 2013 experiment between these two groups did not differ significantly, differences in the unweighted characteristics of the sample are compelling. The increased weights assigned to cellphone sample respondents in the listed landline and RDD cellphone combined sample are also problematic, in part because they produce a higher design effect. The use of an RDD list-assisted landline sample also simplifies the weighting process significantly and reduces potential biases introduced during weighting by the use of the listed landline sample. In addition, the spring 2013 experiment was conducted outside of an election context, meaning the differing sample frames' precise impact on presidential ballot question results among a likely voter sample in an election environment cannot be estimated.

Because of the unweighted differences detected between these samples, and the other considerations discussed above, Gallup has made the decision to return to using an RDD list-assisted landline sample frame for the landline portion of all tracking surveys.

B. SURVEY FIELD MANAGEMENT

3. THE GALLUP NAME

Issue: A possible explanation for a house effect in Gallup data could be some association respondents have with the name "Gallup" when interviewed by Gallup interviewers.

Research conducted: To test the impact of the Gallup name, Gallup conducted an experiment in which respondents were randomly assigned to one of two conditions: a "Gallup" interviewer conducting a "Gallup Poll," and a "Selection Research" interviewer conducting a poll for "Selection Research, Inc." Caller ID information was also altered accordingly to reflect "Gallup" or "Sel. Research" at the time of the call.

Findings: The results showed no difference in the political outcome variables of presidential approval and party identification. The overall response rate was, however, higher with the traditional Gallup identification, which theoretically improved the representativeness of the sample.

Conclusion: The results of this experiment indicate that the Gallup name, by itself, was not a factor that caused Gallup to underestimate Obama's vote share in the 2012 presidential election.

4. RACE OF INTERVIEWER

Issue: Survey researchers have studied the interaction between the race of an interviewer and respondent cooperation and respondent statements for a number of years. Researchers at Stanford University presented analysis at the 2012 American Association of Public Opinion Research (AAPOR) annual conference on the race-of-interviewer effect on survey cooperation. Specifically, researchers found that black interviewers gained marginally better cooperation among black respondents than did white interviewers. A review of 2012 Gallup Daily tracking data showed that black interviewers completed a slightly higher percentage of interviews with black and other nonwhite respondents than did white interviewers. This race-of-interviewer effect on respondent cooperation was small, as is seen in previous research, and there was no significant impact on vote choice.

Research conducted: In order to test whether the proportion of interviews conducted by black interviewers affected its survey results in its 2012 presidential election polling, Gallup researchers simulated an increase in the overall proportion of black interviewers, using Nov. 1-4 pre-election polling data. Gallup then conducted an in-depth experiment in April 2013, using a shadow sample with an interviewing force consisting of 50% black interviewers. The diverse interviewing force and the regular interviewing force (with 10% black interviewers) conducted interviews using the same questionnaire during the same time period. Interviewers in both studies were blind to the purpose of the experiment.

Findings: The interviewing team with the higher percentage of black interviewers conducted only a slightly higher proportion of interviews with black respondents than the regular Gallup interviewing team did, which is consistent with prior research. There were no differences in the results from the two groups in the percentages of interviews completed in the broader non-Hispanic white/all other categories. There were essentially no differences in Obama's unweighted job approval rating between the two research teams. There were slight effects on the unweighted distribution of other political variables, but there were no differences that could be conclusively shown to be due to the increase in the proportion of black interviewers. Black interviewers on the interviewing team with a higher percentage of black interviewers did not gain greater cooperation than white interviewers among black respondents, and there were no differences in cooperation compared with the black interviewers on the regular Gallup interviewing team. Additionally, respondents interviewed by black interviewers on the interviewing team with a higher percentage of black interviewers were not more likely to vote for Obama than those interviewed by white interviewers on the same team.

Additionally, simulations of an increase in black interviewers in the final Nov. 1-4 election poll did not significantly alter the presidential vote results.

Conclusion: The results of simulations on Nov. 1-4 data and the spring 2013 experiment did not show that the racial composition of the interviewing force was related to Gallup's underestimation of Obama's vote share.

5. GENDER OF INTERVIEWER

Issue: The ratio of male and female interviewers working on a survey project could theoretically affect the pattern of responses to certain political questions.

Research conducted: Oct. 1-Nov. 4 election data were reviewed for a possible gender-of-interviewer effect. A simulation was also conducted to test whether changing the gender proportion of the interviewing staff would affect the results.

Findings: During the Oct. 1-Nov. 4 election tracking period, 52% of interviews were conducted by women and 48% were conducted by men. In the final Nov. 1-4 election poll, the gender distribution of interviewers was 53% men and 47% women. There was a slight tendency among respondents interviewed by women during the 2012 election tracking polls to report they were voting for Obama, and for those interviewed by men to report they were voting for Romney. This is not a function of differences in gender of respondents interviewed across gender of interviewer; the data show that female interviewers were actually slightly more likely to interview men across the Oct. 1-Nov. 4 tracking period, and in the Nov. 1-4 tracking period specifically. During the Oct. 1-Nov. 4 time period, male interviewers more frequently interviewed women.

Changes in the gender distribution of interviewers have almost no effect on the ballot outcome, given that the differences in ballot choice between male and female interviewers are so slight. Simulations of increases in the percentage of female interviewers to 60% show a 0.2-point gain for Obama and a 0.1-point loss for Romney in the ballot outcome, neither of which is statistically significant.

The gender composition of the Gallup interviewing force varied across individual days during the 2012 presidential election tracking, but the analysis indicates the variation did not contribute to changes in the daily pattern of vote choice.

Conclusion: The results of these analyses indicate the gender composition of the interviewers working on the 2012 presidential election tracking project did not affect the ballot outcome, although Gallup will continue to conduct research on the most appropriate standards for the gender ratio of interviewers on political projects.

6. NEUTRAL PROBING OF "DON'T KNOW" AND "REFUSED" RESPONSES

Issue: Gallup interviewers are trained to probe for responses when respondents provide initial "don't know" or "undecided" responses, or refuse to answer. Interviewers are trained to do so because it is assumed that respondents who do not initially express a response to a political question, such as the presidential ballot, have inclinations that are reliable and, in the case of the presidential vote, predictive of their actual behavior.

Research conducted: Gallup conducted an experiment to measure the implications of probing on political outcome variables. During a one-week period in April 2013, interviewers recorded each time they probed an initial "don't know," undecided or refusal responses when asking about presidential job approval and party identification.

Findings: The results show that probing decreases the "don't know" and refusal response rates, as expected, but does not alter the overall proportionality of presidential job approval or party identification responses.

Conclusion: The results of this test suggest that interviewer training to probe initial "don't know," undecided, and refusal responses does not affect political outcome variables.

7. GEOGRAPHIC DISTRIBUTION OF INTERVIEWS

Issue: Gallup conducted its 2012 pre-election polling throughout the United States -- across all regions. To control the geographic distribution of interviewing, regional controls mandate that a specific, proportionate number of interviews be conducted in each of four major regions of the country (East, South, Midwest, and West). The interviews completed each night were weighted to targets derived from the Census Bureau for the distribution of the aged 18 and older population across the regions.

The distribution of interviews *within* these regions are not controlled by minimum sample size requirements. Factors relating to the mechanics of interviewing on a daily and nightly basis can affect these within-region distributions, as can factors relating to differential response rates within regions. The factors relating to the mechanics of interviewing primarily revolve around the timing of interviewing within time zone, which is based on sample release and pace of interviewing within each time zone. The South, Midwest, and West regions each include multiple time zones, while the East region is completely in the Eastern time zone. Thus, interviewing is staggered throughout the evening based on the local time within each time zone, and does not start before 5:00 p.m. local time in any time zone. Because there are no minimum sample size requirements by time zone, it is possible that too many interviews within a region can be conducted in that region's "earlier" time zone, and too few in the "later" time zone for that region.

Research conducted: An in-depth analysis shows that the overall proportionality of interviewing across time zones during the Oct. 1-Nov. 4 time period was close to population targets. There was, however, some variation in proportionality of interviews across time zone *within* region, mainly a disproportionately higher number of interviews in the Central time zone in the Midwest and South, and a disproportionately higher number of interviews in the Mountain time zone in the West. These differences are potentially significant because there are differences in election preference by time zone (i.e., specific states within each time zone) within region.

Gallup researchers conducted a simulation in which data were weighted by time zone within region.

Findings: Obama's vote share increased by one percentage point in the final Nov. 1-4 poll after weighting by time zone within region.

State representation: The proportion of interviews conducted within each state is not controlled for in national samples, and can vary within region. As is true for most national samples, larger states are generally underrepresented, while smaller states are overrepresented. The impact of this disproportionate turnout by state was examined using simulations that weighted states proportionally within region.

These simulations showed little change in the overall ballot estimate, although such simulations may not reflect what occurred or might have occurred in the 2012 pre-election environment, and simulations cannot always replicate the actual impact of having more interviews with respondents in specific areas.

Density: As is true with most surveys, interviews are disproportionately conducted in less dense areas because response rates among residents living in highly urban and densely populated areas are lower than those among residents living in suburban and rural areas. This is potentially problematic because of the relationship between population density and political orientation. To address these issues, in its Oct. 1-Nov. 4, 2012, election tracking, Gallup weighted all samples by population density, arraying all counties in the U.S. into five quintiles based on their population density, and weighting the distribution of national interviews to those quintile targets (the target being 20% of completed interviews in each quintile). The weighting brought each quintile in the base underlying sample for the Oct. 1-Nov. 4 surveys close to their correct proportions, although the proportion of interviews in the densest quintile remained marginally below the 20% target in the final Nov. 1-4 sample. Weighting, however, may not alone correct for disproportionality within density quintile in the underlying sample.

Conclusion: Stricter controls on interviewing by time zone within regions would most likely have increased Obama's margin over Romney in Gallup's election tracking by varying amounts at various points in 2012, including by at least one percentage point in the final Nov. 1-4 sample. Gallup is currently working in close consultation with outside experts in the development of a comprehensive plan for appropriate sample allocation controls, including the management of sample by time zones within regions, and the potential management of specific states as their own sample, given the overall difficulty of interviewing in those states.

8. INTERVIEW COMPLETION TIME

Issue: All presidential election tracking interviewing was conducted from interviewing centers located in the Central time zone, with weeknight interviewing conducted between 5 p.m. and 9 p.m. local respondent time (four hours total per time zone per day), Saturday interviewing between 10 a.m. and 3 p.m. local respondent time (five hours total per time zone), and Sunday interviewing between 2 p.m. and 8 p.m. local respondent time (six hours total per time zone). A higher proportion of interviews is conducted weeknights between 4 p.m. and 7 p.m. Central time zone than between 7 p.m. and 10 p.m. Central time zone. Interviewing is also generally conducted earlier in the evening in the Eastern time zone. Interviews in the Mountain time zone in 2012 did not begin until 6 p.m. local time for sample control reasons.

There is a requirement to fulfill the minimum sample size requirements by region for each day or night of interviewing because each day's interviewing is a stand-alone sample with required overall sample size requirements of completed interviews. Interview timing is more problematic on weeknights than on weekends, given the longer interviewing time windows and greater flexibility on weekends to extend interviewing hours if needed to finish interviewing.

Research conducted: Gallup researchers analyzed local interview completion time for potential outcome effects. The distribution of local interview completion times in the final, Nov. 1-4, pre-election poll was fairly balanced, but did vary by region, with the bulk of Eastern interviewing done in the earlier 5-7 p.m. time period and the bulk of Midwestern and Western interviewing done in later time periods. Regional skews in local interview completion time in the yearlong Daily tracking poll were not as large. This is partly because the final poll had only two nights of weeknight interviewing, compared with longer time periods for the Election tracking poll and the underlying Daily tracking poll conducted all year. Across all regions in the final pre-election poll, Obama performed better in calls made earlier on the two weeknights (Nov. 1-2). This relationship is related to the front-loaded Eastern interviewing, as Obama had an expected and sizable advantage in the East that offset Romney's advantages in the other regions on earlier calls. There were no meaningful differences in candidate support by time of call across all regions in the larger aggregate of pre-election polling data, and a small difference in the April-September Daily tracking polling data, which could be a function of delayed calling in the Mountain time zone, which is more Republican.

Findings: These simulations suggest that respondent age, race, and gender are similar by interviewing time. To the extent that there appear to be differences, they appear to be isolated or explained by Gallup strategies for managing completed interviews by time zone and region on weeknights or by other factors in the data for a particular time period.

Conclusion: Aside from a few instances in which early completion of interviewing led to fewer cellphone interviews in the West, it does not appear that calling times made a difference in the accuracy of Gallup's election data. Based on these simulations, it does not appear that distributing the interviews differently across the weeknight interviewing hours would have significantly changed Gallup's final election estimates.

9. CELLPHONE AND LANDLINE PHONE DISTRIBUTION

Issue: The distribution of cellphone and landline interviews during the Oct. 1-Nov. 4 polling period was 50%/50% but varied at times by region.

Research conducted: Gallup analyzed the distribution of cellphone and landline interviews in the context of the 2012 presidential election.

Findings: The distribution of cellphone and landline telephone interviews across regions varied during Gallup's 2012 pre-election polling, and more specifically early on in the pre-election time period. At points in October, the overall minimum sample size requirement for cellphones was disproportionately met with interviews completed in the East, South, and Midwest regions of the country, and fewer cellphone interviews were conducted in the West. Although sample size requirements by region were implemented for the landline telephone interviews throughout the pre-election polling period, regional cellphone minimum sample size requirements were not implemented until the middle of October. In the final Nov. 1-4 poll, research shows that the distribution of completed cellphone and landline telephone interviews was proportional across regions.

Conclusion: The proportion of cellphone and landline telephone interviews completed in the final Nov. 1-4 poll were on target because of the regional minimum sample size controls implemented in October. Thus, based on this analysis, the cellphone and landline distribution did not contribute to Gallup's differing from the final popular vote election result.

C. DATA HANDLING

10. MEASURING AND WEIGHTING RACE

Issue: Gallup weights its national samples according to the U.S. Census Bureau's Current Population Survey (CPS) total population aged 18 and older targets on key demographic variables. Samples in the 2012 presidential election were weighted to correct for unequal selection probability based on the number of adults in the landline household. Samples were also weighted to match national targets of age by gender, race, Hispanic ethnicity, education, region, population density, and phone status.

Race and ethnicity are complex variables to measure. In 2012, Gallup collected race and ethnicity information using a binary "yes" and "no" Hispanic ethnicity question, followed by a series of binary "yes" and "no" questions for each of the Census race categories. Race and ethnicity were then weighted separately to two census categories -- Hispanic, and "black any."

Research conducted: After the 2012 presidential election, Gallup reviewed its method for collecting race and ethnicity information from survey respondents and its overall weighting process.

Gallup reviewed the sample composition of the final Nov. 1-4 dataset, and simulated various weighting approaches to determine if different approaches could have produced different election estimates. These simulations included relaxing and removing the data trimming processes. Data trimming was used in the fall of

2012, meaning weight sizes were capped to avoid significantly increasing the design effect and margin of error when individual respondents were given very large weights.

Findings: Throughout the election polling period, the percentage of multi-race and combination of respondents identifying as white and American Indian and Alaska Native were higher than CPS's estimates for the national adult aged 18 and older population.

The Nov. 1-4 sample neared the demographic characteristics established by CPS after weighting, although the sample slightly underrepresented blacks and Hispanics (Appendix A, Table 1). The proportions of all other respondents were not weighted explicitly, but shrank after race weighting because of the larger weights assigned to Hispanic and black respondents. The sample included an unusually high percentage of nonblack and non-Hispanic respondents who also identified as American Indian or Alaska Native.

Simulations of relaxing and removal of trimming caps produced small differences in the resulting estimates, but at the cost of increasing the design effect.

Conclusion: Gallup concluded that the collection of race information in a series of forced yes/no questions led to over reporting of multi-race and American Indian/Alaska Native groups, which could have produced unknown effects from the race weights. Based on the results of this analysis, Gallup modified its race measurement and weighting procedures. Gallup now reads respondents all race categories and then asks respondents to select up to five race categories from that list. The percentage of respondents selecting American Indian/Alaska Native has decreased since this change was implemented, as has the percentage reporting more than one race, which is now closer to CPS targets.

National adult weighting is now based on four race/ethnicity categories: Hispanic, white non-Hispanic, black non-Hispanic, and other non-Hispanic. In addition to the changes in race measurement and race weighting, other weighting procedures used during Gallup's pre-election polling were revised in 2013, including expanded selection probability and more dual frame compositing. While retroactive simulations of these new weighting procedures using the 2012 data did not show significant changes in the ballot outcome measures, it is not possible to determine their potential impact had they been used during the 2012 pre-election polling, given the 2013 institution of new race measurement procedures.

11. HANDLING OF THIRD-PARTY CANDIDATES

Issue: The ballot used in Gallup's 2012 pre-election surveys included the names of the presidential and vice presidential candidates for the Democratic and Republican parties, along with their party identification. Respondents were allowed to volunteer a third-party candidate, but no third-party candidates' names were read explicitly. Respondents who said they were unsure or undecided were asked which candidate they leaned toward, and respondents who did not lean one way or the other were classified as undecided. All of these reflect Gallup's historical procedures.

Research conducted: Gallup conducted a thorough review of how other major polls asked the ballot question during the 2012 presidential election period.

Findings: Gallup's ballot structure was similar to that of most major polls. Most ballot questions used by major firms did not include an explicit mention of third-party candidates, although at least one firm used a more complex system based on the presence of third-party candidates on individual state ballots. Gallup's final estimate assumed that 3% of the vote share would be assigned to third-party candidates, and the actual result was 2%.

Conclusion: This analysis suggests that the way in which respondents were allowed to indicate a potential vote for a third-party candidate, by itself, was not related to Gallup's underestimation of Obama's vote share during the presidential election.

12. CANDIDATE NAME ORDER IN QUESTION

Issue: Gallup randomly rotates the order in which candidates' names are read on its presidential ballot question because research indicates that ballot order can affect how respondents answer survey items. A review of the questionnaires used by other major survey firms found that all firms randomly rotated the order in which the candidates' names were read in their ballot question.

Research conducted: Gallup conducted a thorough analysis of candidate order and how that order affected voter preference using its Nov. 1-4 election data.

Findings: This analysis demonstrates a primacy effect such that support for each party's presidential and vice-presidential candidates was higher when those names were read first than when they were read second. The order effect was more than twice as large in the final poll Nov. 1-4 than in the entire tracking period. Gallup's review shows that the order in which candidate names were read was approximately 50/50 in both the entire Oct. 1-Nov. 4 tracking period and the final election poll, conducted Nov. 1-4.

Conclusion: This analysis indicates that the randomization of ballot order procedures worked as planned and thus that the effect of the order in which candidate names were read on the phone was equalized across candidates.

Although most major survey firms randomly rotate candidate names on their pre-election polls, the actual order in which the candidates' names appear on ballots varies by state. Thus, for many respondents, the random order they hear on the phone will not match the order in which they will see the candidates' names when they vote. Further, the order of candidates may have a different impact on voters in an election setting -- involving the use of voting machines or paper ballots -- than it does on the phone. Hence, further research is required on this topic, to understand how telephone surveys might better mimic the election experience.

13. LIKELY VOTER ESTIMATING

Issue: In the last month of the 2012 campaign, including the final poll conducted Nov. 1-4, Gallup based its pre-election estimate of the vote on a subsample of likely voters. Gallup's likely voter model involves the use of seven questions to score respondents' likelihood to vote. Gallup uses the results to sort respondents into likely and non-likely voters. The questions cover respondents' intent to vote, interest in the election, past voting behavior, and knowledge of where to vote.

Research conducted: Gallup analyzed the performance of its seven-question likely voter model in the final pre-election poll as well as throughout October, evaluating how Gallup's likely voter numbers differed from its registered voter numbers. Researchers then compared this to the performance of the model in prior election years, as well as to other firms' 2012 likely voter models, to the extent possible. Gallup also reviewed how each question in the model performed, and how various modifications to the model would have affected the final estimate.

Findings: Gallup's likely voter model shifted the race four points in Romney's favor compared with its registered voter estimate, both in the final poll and on average in the last month of the campaign. Comparison to other research firms' models -- to the extent sufficient details on those are available -- reveals that these firms' likely voter screens also pushed the margin in Romney's favor. However, the average shift among these polls was three points, one point less than Gallup's.

Gallup's turnout model is not materially different from most other survey organizations' models, in that it sorts respondents into likely voters and non-likely voters, and bases the final pre-election outcome estimate on the former. Gallup does this by scoring respondents' likelihood of voting according to their answers to seven turnout questions, and then defining all top-scoring respondents (6s and 7s) as likely voters and all others as nonvoters. Some firms also use scoring to identify likely voters, while other firms define likely voters as those

who respond correctly to each of a fixed set of screening questions, and require likely voters to answer all of these "correctly." At least one firm uses a probabilistic weighting method, in which all registered voters who say they plan to vote are included in the likely voter sample to varying degrees, depending on their apparent likelihood to vote. Gallup's process last year produced a likely voter sample that represented a larger proportion of the national adult population (74%) -- essentially a looser screen -- than any other firm's likely voter sample for which we were able to obtain the relevant information.

An analysis of the specific questions that make up various firms' models reveals that Gallup's component questions are more heavily weighted toward past voting, although reducing this emphasis would not by itself have made a significant difference. However, removing other variables from the calculations or modifying the process in other ways, including the removal of the "thought" question, could have changed Gallup's final estimate, as would have other sequences of use of the likely voter questions.

By its nature, research on estimating the likely voter population needs to be conducted in the context of an election environment. Gallup will be conducting such research in the fall in conjunction with the University of Michigan Program in Survey Methodology, using the New Jersey and Virginia gubernatorial elections as real-world laboratories, in order to obtain a clearer picture of how traditional likely voter models perform and how to potentially revise or replace them, particularly in light of changing campaign dynamics and the changing survey environment.

Conclusion: The fundamental assumption justifying likely voter models is that a broad national adult sample, or even registered voter sample, does not reflect the demographic or political profile of the real electorate. As long as the underlying sample is representative of the U.S. public, applying a likely voter model makes sense. Also, historical turnout patterns show that likely voter models should have the effect of increasing support for the Republican candidate. This is how Gallup's model worked in 2012, and apparently how other firms' models worked but to a lesser degree.

Nevertheless, nearly all firms underestimated support for Obama. Gallup's likely voter analysis largely speaks to the difference between Gallup's likely voter model and other firms' models, indicating the need for improvements that might be made based on further research. By making the vote choice less Democratic/more Republican, Gallup's likely voter model performed in the same general way as the models that all other election polling firms used. Gallup's four-point shift toward Romney, however, was greater than other polls' shifts. And, various combinations of the use or weighting of likely voter questions would have changed Gallup's final estimate of the presidential vote to be more in line with other firms'. Thus, Gallup will continue to investigate the model's performance, both with upcoming validation studies of the 2013 gubernatorial elections, and in the context of additional research on the underlying sample. If changes in the research and election environments have shifted national samples closer to likely voters than has been the case historically, then the likely voter model needs to be adapted to correct for it.

As detailed elsewhere in this report, Gallup's underlying sample reflected a house effect that skewed in ways that made it less accurate in estimating the election outcome. If the modifications to Gallup's procedures put in place now and in the future mitigate this house effect, then retaining some modified version of Gallup's traditional likely voter model in future elections may be appropriate. At the same time, planned research will explore the advisability of major revisions or even a replacement of the model, based on the possibility of further declining response rates, increases in early voting and campaign efforts to reach voters, and penetration of new technologies that may change the way researchers reach voters by 2016. Gallup is planning a number of election-related experiments this fall in conjunction with the University of Michigan's Survey Methodology program.

APPENDIX A

Table 1. Race and Ethnicity of Gallup's Final Nov. 1-4 National Adult Sample vs. CPS National Adult Population Targets

	Gallup Nov 1-4 National Adult Sample	CPS National Adult Total Population Targets, March 2011
White non-Hispanic	69.5%	68%
Hispanic	13.2%	14%
Black any	12.0%	N/A
Black alone	11.5%	12%

Table 2. Gender of Gallup's Final Nov. 1-4 National Adult Sample vs. CPS National Adult Population Targets

	Gallup Nov 1-4 National Adult Sample	CPS National Adult Total Population Targets, March 2011
Men	49.0%	49%
Women	51.0%	51%

Table 3. Race and Ethnicity of Gallup's Nov. 1-4 Likely Voter Sample Compared With CPS Estimates of 2013 Election Turnout

	Gallup Nov 1-4 Likely Voters	CPS May 2013 Post-Election Survey	Gallup to CPS May 2013 Post-Election Survey	
	%	%	Pct. pts.	
White Non-Hispanic	76.9	73.7	-3.2	
Hispanic	7.3	8.4	1.1	
Black any	11.4	13.9	2.5	
Black alone	11.1*	13.4	2.3	*Black non-Hispanic
Men	47.8	46.3	-1.5	
Women	52.2	53.7	1.5	
18 to 24	9.0	8.5	-0.5	
25 to 44	30.1	30.0	-0.1	
45 to 64	39.4	39.1	-0.3	
65 to 74	14.5	12.9	-1.6	
75+	7.0	9.4	2.4	
Married, spouse present and absent	65.5	59.3	-6.2	
Widowed	4.7	6.4	1.7	
Divorced	6.6	10.4	3.8	
Separated	1.1	1.7	0.6	
Never married	21.8*	22.0	0.2	*19.5% single, never married + 2.3% domestic partner