The geographic accuracy provided in cellphone samples has been an issue of particular concern for telephone surveys. Surveyors use this information for geographic sample stratification and analysis; however, the location generated using random digit dialing (RDD) methods is typically the rate center in which the phone was activated and not the owner’s current residence. Researchers have found the location is fairly accurate at the regional (census region) level. However, the error rate increases when moving to the state and local levels. Cellphone error rates are higher among ‘cell-only’ (cellphone users who do not have a landline telephone) respondents than for ‘dual-users’ (cellphone users who have a cellphone and landline telephone). Among respondents sampled from the landline frame, the error rate (difference between location on the sample file and current location of residence) is smaller. Past research estimates the state error rate is about 1% for a landline sample and 12% for the ‘cell-only’ group (Skalland and Khare 2013). This paper builds upon previous research by providing state level estimates of inaccuracy for both ‘cell-only’ and ‘dual-users.’ It also examines the trend of state-level inaccuracy estimates over 18 months using data from the Gallup Daily Tracking survey. Finally, we examine the characteristics of ‘cell-only’ and ‘dual-users’ and implications of geographic inaccuracy of cellphone samples on the state level dual-frame RDD estimates.

INTRODUCTION

Researchers frequently employ sample stratification by geographic region when conducting a telephone study using a dual-frame random digit dialing (RDD) design. For some surveys, the population may be restricted to a state, county, or other area. In these studies, researchers must assign each number to a geographic area at the sampling stage. In RDD cellphone samples, the geographic location for a cellphone number is typically based on the cellphone rate center—the location in which the cellphone was purchased. Therefore, the information derived from a cellphone number may be inaccurate if the cellphone is purchased at a location different from where the cellphone owner resides or if the owner moves after the cellphone purchase. These inaccuracies have a negative impact on survey bias, variance, and costs (Skalland and Khare 2013).
Christian et al. (2009) estimated cellphone sample inaccuracy rates were 6% at the census region, 10% at the state level, and 41% at the county level for all adult cellphone users. For cell-only adults, the estimates were 6%, 12% and 43% respectively. Montgomery et al. (2011) inaccuracy rate estimates among all cellphones are 14% at the state and 65.4% at the county level. Benford et al. (2012) reported inaccuracy rates among all cellphone respondents as 5% at the regional, 9% at the state, and 44% at the county level. The cellphone sample inaccuracy rate reported by Skalland and Khare (2013) at the state level for cell-only adults was 11.5%. Dost and McGeeney (2016) estimated cellphone sample geographic inaccuracy rates among all cellphone adults as around 5% at the census region and 10% at the state level. For cell-only adults, the corresponding estimates were 7 and 14%, respectively.

Skalland and Khare (2013) presented overcoverage and undercoverage rates for cell-only adults by each state’s cellphone sampling frame and discussed the implications on bias and variance of dual-frame RDD estimates. Overcoverage was defined as the percentage of cellphone-only adults sampled from the state’s cellphone frame that reside in a different state. Undercoverage was defined as the percentage of cell-only adults that reside in the state whose cellphone number appears on a different state’s cellphone sampling frame. This paper builds upon this research by generating overcoverage and undercoverage estimates by each state’s cellphone sampling frame for ‘cell-only users’ and ‘dual-users,’ and ‘all-cell users’ (‘cell-only’ and ‘dual-users’ combined).

**METHODS**

The data presented in this paper are from three 6-month datasets of the Gallup Daily Tracking survey. The survey includes daily interviews with approximately 500 U.S. adults, and utilizes a dual-frame RDD design. For the landline frame, a list-assisted telephone sampling method and a random, within-household selection method are used. For the cellphone frame, the cellphone numbers are randomly generated among all exchanges dedicated to cellphone numbers.

Table 1 presents time-periods and the corresponding sample sizes.

| Dataset date          | Total completed surveys | Total completed surveys from cellphone sample frame |
|-----------------------|-------------------------|---------------------------------------------------|
| January–June 2017     | 89,792                  | 59,713                                            |
| July–December 2016    | 86,828                  | 49,074                                            |
| January–June 2016     | 90,960                  | 49,540                                            |

The survey weights for each dataset were calculated following Kennedy (2007).

The estimates presented relate to cellphone sampling frames only. While talking about inaccuracy rates for ‘all-cell,’ ‘cell-only,’ or ‘dual-user’ populations, the analysis is restricted to telephone numbers sampled from the
cellphone frames. Completed surveys belonging to those groups that were originally sampled from landline frames are excluded from the inaccuracy rates. The error rate is significantly lower (around 1%) for landline sample frames, and the focus in this paper is to examine the issue of geographic inaccuracy related to cellphone frames only. All estimates presented in this paper are for U.S. adults only.

**ESTIMATES OF CELLPHONE SAMPLE GEOGRAPHIC INACCURACY**

Table 2 presents inaccuracy rates at different geographic levels for different time periods and for 'all-cell,' 'cell-only,' and 'dual-user'. The inaccuracy rates at the census region are relatively low for all time periods and subpopulations. Based on the location designated on the sample file and current residence, the state level error rate, for example, is 12.5% for 'all-cell users' ('cell-only' and 'dual-users' combined), 13.4% for 'cell-only,' and 9% for 'dual-users' for January–June 2017.

| Geographic region | Population | Jan – June 2016 | July – Dec 2016 | Jan – June 2017 |
|-------------------|------------|-----------------|-----------------|-----------------|
| Census region     | All-cell   | 6.1 (5.8, 6.3)  | 6.3 (6.1, 6.6)  | 6.5 (6.3, 6.7)  |
|                   | Cell-only  | 6.7 (6.4, 7.0)  | 7.0 (6.7, 7.3)  | 7.0 (6.8, 7.3)  |
|                   | Dual-users | 4.1 (3.7, 4.4)  | 4.3 (3.9, 4.7)  | 4.2 (3.9, 4.5)  |
| State             | All-cell   | 11.7 (11.4, 12.0)| 12.2 (11.8, 12.5)| 12.5 (12.2, 12.8)|
|                   | Cell-only  | 12.7 (12.3, 13.1)| 13.2 (12.8, 13.6)| 13.4 (13.0, 13.7)|
|                   | Dual-users | 8.7 (8.2, 9.2)  | 9.0 (8.5, 9.5)  | 9.0 (8.5, 9.4)  |
| County            | All-cell   | 39.4 (38.9, 39.9)| 39.1 (38.6, 39.6)| 40.5 (40.0, 40.9)|
|                   | Cell-only  | 39.9 (39.3, 40.5)| 39.7 (39.1, 40.4)| 40.9 (40.4, 41.4)|
|                   | Dual-users | 37.8 (37.0, 38.7)| 37.1 (36.2, 38.0)| 38.7 (37.9, 39.5)|

The inaccuracy rates increase as the area gets smaller. The inaccuracy rates are lower for 'dual-users' as compared with 'cell-only users.' The estimates and corresponding confidence intervals for the other time periods are very similar. The estimates are stable over time with a slight increasing trend. These estimates are consistent with those reported in the past by Christian et al. (2009), Benford et al. (2012), Skalland and Khare (2013) and Dost and McGeeney (2016).

Table 3 presents overcoverage (OC) and undercoverage (UC) estimates by each state’s cellphone sampling frame of 'all-cell users' ('cell-only' and 'dual-users' combined). Following Skalland and Khare (2013), the overcoverage (OC) rate was defined as the percentage of 'all-cell users' sampled from the state’s cellphone frame that reside in a different state. The undercoverage (UC) rate was defined as the percentage of all cell users that reside in the state whose cellphone number appears on a different state’s cellphone sampling frame.
| State           | Jan-Jun 2016 | Jul-Dec 2016 | Jan-Jun 2017 |
|-----------------|--------------|--------------|--------------|
| National        | 11.7(11.4,12.0) | 12.2(11.8,12.5) | 12.5(12.2,12.8) |
| Alabama+        | 8.6(6.5,11.3)  | 12.0(9.4,15.2)  | 9.4(7.4,11.8)  |
| Alaska          | 18.9(12.1,28.3) | 20.0(12.3,30.9) | 18.3(11.7,27.6) |
| Arizona         | 10.0(8.1,12.3)  | 9.2(7.4,11.4)  | 11.3(9.5,13.3)  |
| Arkansas        | 11.7(8.8,15.4)  | 10.1(7.2,14.0)  | 11.8(8.6,15.9)  |
| California+@    | 7.3(6.6,8.1)   | 6.7(6.0,7.5)   | 6.9(6.2,7.7)   |
| Colorado        | 16.1(13.5,19.2) | 17.0(14.2,20.2) | 19.9(17.2,22.9) |
| Connecticut     | 19.0(15.7,22.9) | 19.3(15.9,23.2) | 16.8(13.5,20.8) |
| Delaware#       | 22.6(15.1,32.5) | 19.8(12.8,29.2) | 18.8(11.6,29.0) |
| District Of Columbia*# | 58.5(50.3,66.2) | 45.4(36.9,54.1) | 61.1(52.8,68.8) |
| Florida+        | 8.8(7.8,9.9)   | 13.1(11.9,14.4) | 8.7(7.6,9.8)   |
| Georgia         | 10.4(8.9,12.1)  | 12.7(10.9,14.7) | 12.3(10.6,14.2) |
| Hawaii#         | 14.8(10.1,21.0) | 10.6(8.6,16.2)  | 26.5(18.8,36.0) |
| Idaho           | 18.5(13.7,24.3) | 19.4(14.3,26.0) | 13.0(9.3,17.8)  |
| Illinois        | 9.8(8.5,11.4)   | 10.3(8.8,12.1)  | 11.2(9.7,13.0)  |
| Indiana         | 11.5(9.4,13.9)  | 9.4(7.7,11.5)   | 12.3(10.1,15.1) |
| Iowa            | 12.3(9.3,16.1)  | 13.5(10.5,17.2) | 17.9(11.5,18.6) |
| Kansas          | 12.1(9.3,15.5)  | 16.1(12.6,20.3) | 17.0(13.4,21.3) |
| Kentucky        | 15.0(12.1,18.3) | 13.1(10.4,16.3) | 16.7(13.5,20.5) |
| Louisiana@      | 15.2(12.3,18.6) | 9.1(6.9,11.8)   | 8.3(6.3,11.0)   |
| Maine@          | 17.1(12.1,23.6) | 13.1(8.9,19.0)  | 15.1(10.7,21.1) |
| Maryland#       | 15.7(13.2,18.6) | 22.3(19.5,25.4) | 19.9(17.1,23.0) |
| Massachusetts   | 16.8(14.5,19.4) | 14.4(12.2,17.0) | 17.8(15.4,20.5) |
| Michigan+@      | 9.1(7.7,10.8)   | 6.0(4.8,7.5)    | 9.8(8.3,11.6)   |
| Minnesota       | 10.6(8.6,13.1)  | 8.3(6.3,10.8)   | 12.7(10.2,15.6) |
| Mississippi+    | 8.6(6.1,11.8)   | 12.3(9.3,16.1)  | 12.2(8.9,16.3)  |
| Missouri        | 10.9(8.8,13.4)  | 12.0(9.9,14.5)  | 12.1(9.9,14.7)  |
| Montana         | 16.7(11.0,24.7) | 13.1(8.8,19.1)  | 21.3(15.1,29.1) |
| Nebraska        | 15.0(11.4,19.7) | 13.5(10.2,17.8) | 18.2(13.9,23.5) |
| Nevada          | 14.2(10.8,18.3) | 20.1(16.1,24.8) | 13.8(10.3,18.3) |
| New Hampshire*  | 23.5(17.7,30.6) | 19.1(13.1,26.9) | 17.4(12.4,23.7) |
| New Jersey      | 16.3(14.3,18.5) | 11.8(10.1,13.7) | 17.8(15.5,20.3) |
| New Mexico      | 14.5(10.7,19.4) | 11.5(8.6,15.3)  | 14.5(9.8,20.8)  |
| New York@       | 16.2(14.9,17.7) | 8.9(7.9,10.0)   | 18.6(17.1,20.3) |
| North Carolina  | 11.5(9.7,13.7)  | 14.9(13.0,16.9) | 10.9(9.2,12.9)  |
| North Dakota#   | 19.8(13.0,29.0) | 26.2(18.4,36.0) | 20.3(13.3,30.0) |
| Ohio@           | 10.9(9.4,12.6)  | 8.4(7.1,10.0)   | 10.9(9.3,12.6)  |
| Oklahoma        | 8.8(6.6,11.6)   | 11.8(8.8,13.9)  | 10.0(7.4,13.4)  |
| Oregon          | 11.4(9.1,14.3)  | 16.7(13.7,20.1) | 14.1(11.2,17.7) |
| Pennsylvania    | 15.4(13.7,17.3) | 10.0(8.6,11.5)  | 14.6(12.8,16.6) |
| Rhode Island*   | 22.0(15.4,30.5) | 21.5(15.2,29.5) | 20.9(14.7,28.8) |
| South Carolina  | 11.2(8.8,14.1)  | 16.9(14.2,20.0) | 11.5(8.9,14.8)  |
| South Dakota+   | 7.9(4.5,13.4)   | 13.3(8.2,20.6)  | 14.2(9.1,21.5)  |

**Survey Practice**

4

Table 3. Estimated overcoverage (OC) and undercoverage (UC) rates of all cell users by state cellphone sampling frame.
The overcoverage and the undercoverage percentages vary by state. High overcoverage (20% or more) are found in smaller states and small overcoverage (10% or less) are found in relatively bigger states. Relatively higher (lower) undercoverage are observed for larger states. The overcoverage and undercoverage percentages did not change much over time except for a few smaller states.

Table 4 presents the overcoverage (OC) and undercoverage (UC) rates of ‘all-cell,’ ‘cell-only,’ and ‘dual-users’ separately for the latest time period (January–June 2017).

| State          | OC (Lower,Upper) | UC (Lower,Upper) |
|----------------|------------------|------------------|
| Tennessee      | 14.0(11.8,16.6)  | 11.1(9.2,13.4)   |
| Texas+@        | 6.4(5.7,7.2)     | 8.6(7.7,9.5)     |
| Utah           | 14.5(11.1,18.7)  | 11.9(9.3,15.0)   |
| Vermont*#      | 40.7(29.1,53.4)  | 24.2(15.7,35.4)  |
| Virginia#      | 15.2(13.2,17.5)  | 18.7(16.4,21.2)  |
| Washington     | 12.5(10.5,14.8)  | 16.8(14.5,19.4)  |
| West Virginia  | 13.5(9.2,19.4)   | 12.9(8.7,18.9)   |
| Wisconsin      | 13.1(10.8,15.9)  | 12.2(9.9,14.9)   |
| Wyoming#       | 21.3(14.2,30.9)  | 14.2(9.3,21.1)   |

*Overcoverage of 20%+ in 2 or more periods
+Overcoverage of <=10% in 2 or more periods
(#)Undercoverage of 20%+ in 2 or more periods
@Undercoverage of <=10% in 2 or more periods
Table 4. Estimated overcoverage (OC) and undercoverage (UC) rates of ‘all-cell’, ‘cell-only’ and ‘dual’ users by state cellphone sampling frame – time period (January–June 2017).

| State              | OC          | All-Cell  |
|--------------------|-------------|-----------|
| National           | 12.5(12.2,12.8) | 12.5(12.2,12.8) |
| Alabama            | 9.4(7.4,11.8)  | 11.9(9.8,14.4)  |
| Alaska             | 18.3(17.2,27.6) | 18.2(12.8,25.4) |
| Arizona            | 10.0(8.3,12.1)  | 15.7(13.7,18.0)  |
| Arkansas           | 11.8(8.5,14.4)  | 14.4(14.4,17.9)  |
| California         | 8.7(8.0,9.5)   | 6.8(6.2,7.4)    |
| Colorado           | 13.8(11.7,16.1) | 23.6(20.9,26.5) |
| Connecticut        | 21.0(17.7,24.7) | 14.4(11.6,17.7) |
| Delaware           | 18.6(18.2,26.3) | 26.0(18.5,35.3) |
| District Of Columbia | 59.7(52.8,66.2) | 42.5(35.8,49.5) |
| Florida            | 10.6(9.6,11.7)  | 14.3(13.2,15.4)  |
| Georgia            | 13.7(12.0,15.5) | 14.7(13.0,16.5)  |
| Hawaii             | 16.3(12.2,21.5) | 20.4(15.0,27.1)  |
| Idaho              | 13.0(9.3,17.8)  | 20.3(15.9,25.5)  |
| Illinois           | 11.4(10.1,12.9) | 11.7(10.2,13.4)  |
| Indiana            | 9.7(7.9,11.8)   | 11.3(9.3,13.5)   |
| Iowa               | 14.3(11.5,17.7) | 12.4(9.9,15.5)   |
| Kansas             | 13.2(10.5,16.6) | 17.1(14.2,20.6)  |
| Kentucky           | 13.6(11.0,16.6) | 12.5(10.1,15.6)  |
| Louisiana          | 9.6(7.6,12.0)   | 8.4(6.5,10.9)    |
| Maine              | 18.0(13.4,23.8) | 7.5(4.8,11.5)    |
| Maryland           | 19.4(16.8,22.2) | 23.0(20.2,26.0)  |
| Massachusetts      | 17.6(15.5,19.8) | 14.6(12.6,16.8)  |
| Michigan           | 10.7(9.3,12.4)  | 7.7(6.4,9.2)     |
| Minnesota          | 12.5(10.4,14.9) | 12.4(10.2,15.0)  |
| Mississippi        | 9.5(7.0,12.7)   | 16.5(13.1,20.7)  |
| Missouri           | 14.3(12.2,16.6) | 12.6(10.6,14.9)  |
| Montana            | 17.1(12.5,23.0) | 21.2(15.7,28.0)  |
| Nebraska           | 16.2(12.6,20.5) | 11.3(8.7,14.5)   |
| Nevada             | 15.8(12.4,19.9) | 17.2(14.0,20.9)  |
| New Hampshire      | 20.5(15.4,26.8) | 18.2(13.7,23.8)  |
| New Jersey         | 18.0(16.1,20.0) | 13.1(11.7,15.2)  |
| New Mexico         | 12.4(9.4,16.1)  | 15.9(12.2,20.5)  |
| New York           | 17.0(15.8,18.4) | 9.9(8.9,11.0)    |
| North Carolina     | 11.6(10.1,13.4) | 15.1(13.4,17.1)  |
| North Dakota       | 24.9(18.2,33.2) | 27.4(20.6,35.5)  |
| Ohio               | 10.8(9.4,12.4)  | 8.6(7.4,10.1)    |
| Oklahoma           | 10.6(8.4,13.3)  | 11.3(8.9,14.3)   |
| Oregon             | 16.4(13.7,19.6) | 14.3(12.1,16.8)  |
| Pennsylvania       | 14.6(13.1,16.2) | 10.3(9.0,11.7)   |
| Rhode Island       | 18.2(13.0,24.9) | 15.2(10.2,22.2)  |
| South Carolina     | 10.1(8.1,12.6)  | 15.7(13.3,18.5)  |
| South Dakota       | 9.9(5.9,16.1)   | 14.9(10.1,21.5)  |
| Tennessee          | 15.4(13.2,17.9) | 13.4(11.4,15.6)  |

Survey Practice 6
The overcoverage and undercoverage percentages for all three populations vary by state. For 'all-cell' and 'cell-only' users, the overcoverage is lowest for Texas (7.2% and 7.3%, respectively) and highest for the District of Columbia (DC) (59.7% and 59.5%, respectively). For 'dual users,' it was lowest for Maine (3.9%) and highest for DC (60.4%). The undercoverage is lowest for California for all three populations. For the 'all-cell' and 'cell-only' users, the undercoverage is highest for DC (42.5% and 47.1%, respectively). For 'dual-users,' the undercoverage is highest for Delaware (22.8%). The smaller states had relatively higher overcoverage and undercoverage percentages. For 'dual-users,' the overcoverage and the undercoverage percentages are generally lower compared to 'all-cell' and 'cell-only' populations. The confidence intervals associated with 'cell-only' estimates are generally narrower than those for 'dual-users' because of larger sample sizes and are asymmetric because CROSSTAB, the SUDAAN procedure employed to construct the confidence intervals, uses logit transformation.

**CHARACTERISTICS OF RESPONDENTS BY MATCH STATUS**

Table 5 presents selected characteristics of 'all-cell users' for January–June 2017 by match status. A match happens when the sampling state is the same as the state of residence. The bias, if any, due to undercoverage is a function of the difference between matched and unmatched groups and the amount of undercoverage. The bivariate relationship between match status and demographic variables was significant for all variables (based on the chi-square test), except gender, indicating a risk of potential bias. Those with a nonmatch status are more likely to be younger adults, non-Hispanic, more educated, belong to households without children, and residents of metropolitan statistical areas (MSAs). Nonmatches are also more common in Southern states as compared with states in Midwest region.
### Table 5. Demographic distributions of adults for ‘all-cell users’ by cellphone sampling state geographic accuracy.

| Demographic variables               | Match (sampling state matches self-reported state of residence) | Nonmatch (sampled state does not match self-reported state of residence) | All-Cell |
|-------------------------------------|-----------------------------------------------------------------|--------------------------------------------------------------------------|---------|
| **Gender**                          |                                                                 |                                                                          |         |
| Male                                | 53.0 (52.4, 53.5)                                               | 52.4 (51.1, 53.6)                                                        | 52.9 (52.4, 53.3) |
| Female                              | 47.0 (46.5, 47.6)                                               | 47.6 (46.4, 48.9)                                                        | 47.1 (46.7, 47.6) |
| Chi-Square P-Value = 0.4111         |                                                                 |                                                                          |         |
| **Age**                             |                                                                |                                                                          |         |
| 18-29                               | 28.3 (27.8, 28.8)                                               | 35.1 (33.9, 36.4)                                                        | 29.1 (28.7, 29.6) |
| 30-49                               | 38.0 (37.5, 38.5)                                               | 41.0 (39.8, 42.2)                                                        | 38.4 (37.9, 38.8) |
| 50+                                 | 33.7 (33.3, 34.2)                                               | 23.9 (23.0, 24.9)                                                        | 32.5 (32.1, 32.9) |
| Chi-Square P-Value = 0.0000         |                                                                 |                                                                          |         |
| **Race/ethnicity**                  |                                                                |                                                                          |         |
| Hispanic                            | 20.7 (20.2, 21.1)                                               | 11.7 (10.9, 12.6)                                                        | 19.6 (19.2, 20.0) |
| NonHisp white only                  | 55.6 (55.1, 56.1)                                               | 64.3 (63.1, 65.5)                                                        | 56.7 (56.2, 57.2) |
| NonHisp black only                  | 11.7 (11.4, 12.1)                                               | 11.1 (10.3, 11.9)                                                        | 11.7 (11.3, 12.0) |
| NonHisp other                       | 12.0 (11.7, 12.3)                                               | 12.9 (12.0, 13.7)                                                        | 12.1 (11.8, 12.4) |
| Chi-Square P-Value = 0.0000         |                                                                 |                                                                          |         |
| **Education**                       |                                                                |                                                                          |         |
| Less than H.S. degree               | 13.5 (13.0, 13.9)                                               | 4.4 (3.8, 5.2)                                                          | 12.3 (12.0, 12.7) |
| High School degree or GED           | 30.4 (30.0, 30.9)                                               | 20.6 (19.5, 21.7)                                                        | 29.2 (28.8, 29.7) |
| Some college, no degree/associate degree | 29.0 (28.5, 29.4)                                           | 25.9 (24.8, 27.0)                                                        | 28.6 (28.2, 29.0) |
| Bachelor or higher                  | 27.1 (26.8, 27.5)                                               | 49.1 (47.9, 50.4)                                                        | 29.9 (29.5, 30.3) |
| Chi-Square P-Value = 0.0000         |                                                                 |                                                                          |         |
| **Marital status**                  |                                                                |                                                                          |         |
| Never married                       | 41.9 (41.4, 42.4)                                               | 43.9 (42.6, 45.1)                                                        | 42.1 (41.7, 42.6) |
| Married                             | 45.1 (44.6, 45.6)                                               | 46.3 (45.1, 47.5)                                                        | 45.3 (44.8, 45.8) |
| Divorce/widowed/separated           | 13.0 (12.7, 13.3)                                               | 9.8 (9.1, 10.6)                                                          | 12.6 (12.3, 12.9) |
| Chi-Square P-Value = 0.0000         |                                                                 |                                                                          |         |
| **Number children in household**    |                                                                |                                                                          |         |
| 0 children                          | 72.0 (71.6, 72.5)                                               | 77.5 (76.5, 78.6)                                                        | 72.7 (72.3, 73.1) |
| 1 child                             | 5.2 (5.0, 5.5)                                                  | 2.9 (2.5, 3.4)                                                          | 4.9 (4.7, 5.1)   |
| 2+ children                         | 22.8 (22.3, 23.2)                                               | 19.5 (18.6, 20.5)                                                        | 22.3 (22.0, 22.8) |
| Chi-Square P-Value = 0.0000         |                                                                 |                                                                          |         |
| **MSA status**                      |                                                                |                                                                          |         |
| In MSA                              | 86.4 (86.1, 86.7)                                               | 89.4 (88.6, 90.1)                                                        | 86.8 (86.5, 87.1) |
| Not in MSA                          | 13.6 (13.3, 13.9)                                               | 10.6 (9.9, 11.4)                                                        | 13.2 (12.9, 13.5) |
| Chi-Square P-Value = 0.0000         |                                                                 |                                                                          |         |
| **Region**                          |                                                                |                                                                          |         |
| Northeast                           | 15.8 (15.6, 16.0)                                               | 14.7 (13.9, 15.5)                                                        | 15.7 (15.5, 15.8) |
| Midwest                             | 20.7 (20.5, 21.0)                                               | 18.2 (17.2, 19.2)                                                        | 20.4 (20.2, 20.6) |
| South                               | 38.3 (38.0, 38.6)                                               | 43.4 (42.2, 44.7)                                                        | 38.9 (38.7, 39.2) |
| West                                | 25.2 (24.9, 25.4)                                               | 23.7 (22.7, 24.8)                                                        | 25.0 (24.7, 25.2) |
| Chi-Square P-Value = 0.0000         |                                                                 |                                                                          |         |

A logistic regression model was also fit using the sample match status as a
0-1 binary dependent variable and the demographic variables in Table 5 as independent variables. The goal was to see if any of these bivariate relationships are nonsignificant after controlling for the other variables. The regression coefficients are presented in Table 6 for ‘all-cell,’ ‘cell-only,’ and ‘dual-users.’

Table 6. Parameter estimates for logistic regression model for ‘all-cell,’ ‘cell-only,’ and ‘dual-users.’

| Effects                  | All-Cell Beta Coeff. | SE Beta | Cell-Only Beta Coeff. | SE Beta | Dual-Users Beta Coeff. | SE Beta |
|--------------------------|----------------------|---------|-----------------------|---------|------------------------|---------|
| Intercept                | -2.24***             | 0.09    | -2.18***              | 0.1     | -2.26***               | 0.18    |
| Gender                   |                      |         |                       |         |                        |         |
| Male                     | 0.04                 | 0.03    | 0.03                  | 0.03    | 0.02                   | 0.06    |
| Female                   | Reference            |         |                       |         |                        |         |
| Age                      |                      |         |                       |         |                        |         |
| 18-29                    | 0.77***              | 0.04    | 0.70***               | 0.05    | 0.26**                 | 0.12    |
| 30-49                    | 0.50***              | 0.04    | 0.41***               | 0.04    | 0.42***                | 0.07    |
| 50+                      | Reference            |         |                       |         |                        |         |
| Race/ethnicity           |                      |         |                       |         |                        |         |
| Hispanic                 | -0.51***             | 0.06    | -0.53***              | 0.07    | -0.34**                | 0.15    |
| NonHisp white            | 0.03                 | 0.04    | 0.07                  | 0.05    | -0.12                  | 0.1     |
| NonHisp black            | -0.11                | 0.06    | -0.09                 | 0.07    | -0.08                  | 0.13    |
| NonHisp other            | Reference            |         |                       |         |                        |         |
| Education                |                      |         |                       |         |                        |         |
| Less than H.S. degree    | -1.45***             | 0.09    | -1.47***              | 0.09    | -1.82***               | 0.3     |
| High School degree or GED| -0.96***             | 0.04    | -1.03***              | 0.04    | -0.63***               | 0.09    |
| Some college, no degree/associate degree | -0.69*** | 0.03    | -0.74***              | 0.04    | -0.43***               | 0.07    |
| Bachelor or higher       | Reference            |         |                       |         |                        |         |
| Marital Status           |                      |         |                       |         |                        |         |
| Never married            | -0.08                | 0.05    | -0.05                 | 0.06    | -0.17                  | 0.12    |
| Married                  | 0.01                 | 0.05    | 0.05                  | 0.06    | -0.03                  | 0.09    |
| Divorce/widowed/separated| Reference            |         |                       |         |                        |         |
| Number Children in Household |                |         |                       |         |                        |         |
| 0 children               | 0.30***              | 0.04    | 0.32***               | 0.04    | 0.08                   | 0.09    |
| 1 child                  | -0.20**              | 0.09    | -0.20**               | 0.1     | -0.25*                 | 0.14    |
| 2+ children              | Reference            |         |                       |         |                        |         |
| MSA status               |                      |         |                       |         |                        |         |
| In MSA                   | 0.12***              | 0.04    | 0.15***               | 0.05    | -0.02                  | 0.09    |
| Not in MSA               | Reference            |         |                       |         |                        |         |
| Region                   |                      |         |                       |         |                        |         |
| Northeast                | 0.03                 | 0.05    | 0.14***               | 0.05    | 0.08                   | 0.1     |
| Midwest                  | Reference            |         |                       |         |                        |         |
| South                    | 0.37***              | 0.04    | 0.34***               | 0.04    | 0.6***                 | 0.08    |
| West                     | 0.15***              | 0.04    | 0.21***               | 0.05    | -0.05                  | 0.1     |

*p<0.10,**p<0.05,***p<0.01
All demographic variables included in the model (Table 6) were statistically significant with the exception of gender for both 'all-cell' and 'cell-only' user populations. For 'dual-users,' the variables marital status, race/ethnicity, and in MSA were nonsignificant. The differences in the demographics of matches and nonmatches were more pronounced for the 'cell-only users' as compared with 'dual-users.' These findings are generally consistent with those of Christian et al. (2009) and Skalland and Khare (2013). The analysis included large sample sizes which may be why more demographic variables were statistically significant. The analysis was carried out using the software SUDAAN (https://www.rti.org/impact/sudaan-statistical-software-analyzing-correlated-data) to account for the complex sample design and the resulting weights. SUDAAN functions CROSSTAB and RLOGIST were used to run statistical tests and fit logistic regression models.

For 'all-cell users,' the response categories of selected survey questions are included in Table 7 along with estimates and corresponding 95% confidence intervals. The differences in the distribution of these variables between matches and nonmatches were statistically significant at 1%. When these variables are added to the logistic regression model (of Table 6) as independent variables, all variables except gender, marital status, MSA, and employment status are significant.

The findings from the same analyses for 'cell-only' and 'dual-users' were similar.
Table 7. Survey estimates for ‘all-cell users’ by cellphone sampling state geographical accuracy.

| Selected Questions          | Match (sampling state matches self-reported state of residence) | Nonmatch (sampled state does not match self-reported state of residence) | All-Cell |
|-----------------------------|-----------------------------------------------------------------|---------------------------------------------------------------------|---------|
| **Religion**                |                                                                 |                                                                    |         |
| important                   | 60.8(60.3,61.3)                                                  | 50.7(49.4,52.0)                                                      | 59.6(59.1,60.0) |
| not important               | 39.2(38.7,39.7)                                                  | 49.3(48.0,50.6)                                                      | 40.4(40.0,40.9) |
| **Party Affiliation**       |                                                                 |                                                                    |         |
| Republican                  | 26.7(26.2,27.1)                                                  | 23.8(22.7,24.8)                                                      | 26.3(25.9,26.7) |
| Democrat                    | 31.8(31.3,32.3)                                                  | 35.3(34.1,36.5)                                                      | 32.3(31.8,32.7) |
| Independent                 | 41.5(41.0,42.0)                                                  | 40.9(39.7,42.2)                                                      | 41.4(41.0,41.9) |
| **Military service**        |                                                                 |                                                                    |         |
| Resp/other family member    | 9.5(9.3,9.8)                                                    | 14.1(13.3,14.9)                                                     | 10.1(9.9,10.4) |
| Other family member         | 10.6(10.3,10.9)                                                 | 11.3(10.6,12.2)                                                     | 10.7(10.4,11.0) |
| No                          | 79.9(79.5,80.2)                                                 | 74.6(73.5,75.6)                                                     | 79.2(78.8,79.6) |
| **Presidential approval**   |                                                                  |                                                                    |         |
| Approve                     | 41.4(40.8,41.9)                                                 | 34.6(33.3,35.9)                                                     | 40.5(39.9,41.0) |
| Disapprove                  | 58.6(58.1,59.2)                                                 | 65.4(64.1,66.7)                                                     | 59.5(59.0,60.1) |
| **Employment**              |                                                                  |                                                                    |         |
| FT (employer)               | 68.5(68.0,69.0)                                                 | 73.3(72.0,74.5)                                                     | 69.1(68.6,69.6) |
| FT (self)                   | 9.2(8.9,9.5)                                                   | 7.5(6.8,8.3)                                                        | 9.0(8.7,9.3)  |
| PT (do not want FT)         | 8.3(8.0,8.6)                                                   | 7.4(6.7,8.2)                                                        | 8.2(7.9,8.5)  |
| Unemployed                  | 5.8(5.5,6.1)                                                   | 4.5(3.9,5.2)                                                        | 5.6(5.3,5.9)  |
| PT (want FT)                | 8.2(7.9,8.6)                                                   | 7.3(6.6,8.1)                                                        | 8.1(7.8,8.4)  |

Chi-Square P-Value=0.0000

**IMPLICATIONS FOR DUAL-FRAME STATE SURVEYS**

For single state dual-frame telephone surveys, the geographic inaccuracy of cellphone samples has operational, cost, and bias implications (Skalland and Khare 2013). For a state survey, cellphones are sampled from that state’s cellphone frame only. The resulting undercoverage of residents living in that state who are present on another state’s frame may introduce bias in estimates. Residents whose cellphone number appears on the state’s cellphone frame but actually live in a different state (overcoverage) may be screened out. Therefore, the overcoverage will not affect estimates but will negatively impact cost.

Following Skalland and Khare (2013), alternative state level estimates were derived using a modified sample (all landline sample respondents and only those cellphone sample respondents whose self-reported state matched the sampling state). Those estimates were compared with the full sample estimates for each state to get a measure of potential noncoverage bias due to geographic inaccuracy of cellphone samples.
Each of the variables were coded 0-1 and two estimates (based on full sample and modified sample) were derived for the mean (or proportion) of each variable. Table 8 presents, for each state, the difference between the estimates. None of these differences were found statistically significant (at 5% level of significance). However, some differences were relatively large, suggesting potential noncoverage bias due to geographic inaccuracy of cellphone samples.

Geographic Inaccuracy of Cellphone Samples

Survey Practice
Table 8. Change in full sample estimates by using matched cellphone interviews only—by true state of residence.

| Self-report state of residence | Religion importance | Trump approval | Fulltime employed |
|-------------------------------|--------------------|---------------|------------------|
| Alabama                       | 0.43               | 0.28          | -0.20            |
| Alaska                        | 0.65               | 1.16          | -3.89*           |
| Arizona                       | 0.15               | 0.26          | 0.07             |
| Arkansas                      | 1.50               | -0.01         | 1.36             |
| California                    | 0.41               | 0.22          | -0.39            |
| Colorado                      | 1.29               | 1.09          | -0.61            |
| Connecticut                   | -0.02              | -0.17         | 0.11             |
| Delaware                      | 2.13*              | 1.71          | -0.07            |
| District of Columbia          | 3.53*              | -0.41         | 0.43             |
| Florida                       | 0.35               | 0.19          | -0.14            |
| Georgia                       | 0.98               | -0.20         | 0.80             |
| Hawaii                        | 1.71               | -1.10         | 0.00             |
| Idaho                         | 1.07               | 0.23          | -1.57            |
| Illinois                      | 0.36               | 0.37          | -0.12            |
| Indiana                       | 0.28               | -0.18         | -0.22            |
| Iowa                          | 0.35               | -0.11         | 0.80             |
| Kansas                        | -1.10              | 0.15          | -1.24            |
| Kentucky                      | -0.48              | 0.16          | -0.46            |
| Louisiana                     | 0.56               | 0.34          | 0.10             |
| Maine                         | 0.07               | -0.35         | 0.89             |
| Maryland                      | 0.24               | 1.57          | -1.12            |
| Massachusetts                 | 0.42               | 0.73          | -0.40            |
| Michigan                      | 0.49               | 0.23          | 0.36             |
| Minnesota                     | 0.27               | 0.44          | -0.30            |
| Mississippi                   | 0.69               | 1.71          | -0.37            |
| Missouri                      | 0.66               | 0.52          | 0.11             |
| Montana                       | -0.08              | -0.05         | 2.28*            |
| Nebraska                      | 0.79               | 1.92          | 0.33             |
| Nevada                        | 0.95               | -1.40         | -0.75            |
| New Hampshire                 | 0.79               | 0.63          | -0.78            |
| New Jersey                    | 0.91               | 0.00          | -0.92            |
| New Mexico                    | -0.48              | 1.13          | -0.78            |
| New York                      | 0.45               | 0.57          | -0.46            |
| North Carolina                | 1.06               | -0.03         | -0.05            |
| North Dakota                  | -2.59*             | -1.54         | 1.32             |
| Ohio                          | -0.47              | 0.17          | 0.03             |
| Oklahoma                      | 0.01               | 0.44          | -0.75            |
| Oregon                        | 0.75               | 0.81          | -0.99            |
| Pennsylvania                  | 0.57               | 0.61          | 0.08             |
| Rhode Island                  | 0.90               | 3.01*         | -0.71            |
| South Carolina                | 0.70               | 0.52          | 0.22             |
| South Dakota                  | 0.74               | -1.23         | 0.02             |
| Tennessee                     | 1.16               | 0.93          | 0.32             |
| Texas                         | 0.63               | 0.26          | 0.03             |
| Utah                          | -1.47              | 0.47          | 0.15             |
| Vermont                       | 1.29               | 0.82          | 1.19             |
| Virginia                      | 0.55               | -0.16         | -0.12            |
| Washington                    | 0.75               | -0.02         | -1.04            |
Table 1. Geographic Inaccuracy of Cellphone Samples

| State     | Inaccuracy Rate | Undercoverage Rate | Overcoverage Rate |
|-----------|-----------------|--------------------|------------------|
| West Virginia | 0.32            | 1.25               | 1.47             |
| Wisconsin  | 0.76            | 0.53               | -0.05            |
| Wyoming    | 2.48*           | 2.22*              | 2.57*            |

*The absolute value is greater than 2.00.

**DISCUSSION**

The mobile nature of cellphones poses a significant problem for geographic sampling, and particularly for small areas. For landline numbers, the inaccuracy rate (around 1%) is considerably smaller. The estimates presented in this paper show that about 12.5% of U.S. cell users do not live in their sampling state and that there is variation across states. For ‘cell-only’ users, the corresponding estimate is 13.4%, whereas it is 9% for ‘dual-users’. These estimates increase over time except for Midwest region where they are generally stable. The inaccuracy rates for ‘dual-users’ is lower than for ‘cell-only users’. Given the percentage of ‘cell-only’ population is growing within cell users, the cellphone sample geographic inaccuracy will continue to be a concern for all areas and in particular for smaller geographic levels.

For a given area, respondents may be identified and screened out of the survey if they do not live there, but eligible respondents who live in the area with cellphone numbers from a different area will not be covered. Screening out respondents from other areas will not cause coverage problems but will lead to higher costs. The undercoverage will result in noncoverage bias. If most of those individuals with inaccurate sample states live in contiguous states, sampling from adjacent states may reduce undercoverage. However, that will increase the overcoverage percentage significantly. For example, if Virginia and DC cellphone sampling frames are also used to sample for Maryland, the undercoverage percentage for all Maryland users will drop from 21.5 to 14.1%. However, the overcoverage will increase from 19.3 to 66.3%, leading to significantly higher survey costs and lead to inefficient use of all resources including time and people. Sampling from selected counties and zip codes from adjoining states may be more effective.

Efforts are being made to improve sample accuracy. Sample providers can now append the billing zip code of the cellphone owner’s address, although this is available for only about 60% of all cellphone numbers (Dutwin 2014). Moreover, this information is appended after the sample is drawn and cannot be used while sampling to improve sample efficiency and reduce sample costs. Address based sampling (ABS) is another alternative to RDD telephone surveys. However, for conducting telephone surveys starting with an ABS sample, telephone numbers will have to be appended to sampled addresses and only about 50–60% of the addresses can be appended with telephone numbers at the national level. Another option to improve estimates may be to employ additional weighting steps to account for correlates of mobility,
the characteristic that may be most related to sampling inaccuracy. Additional research is needed to evaluate if any such strategy will be effective for specific states. Finally, this problem of geographic inaccuracy may worsen over time as more and more households switch from landline to cell phones. The impact of geographic inaccuracy will depend on several factors including the geographic level, the target population, and the topic of the survey.

**AUTHOR INFORMATION**
Stephanie Marken, 901 F Street NW Suite 400, Washington, DC 20004, stephanie_marken@gallup.com, 202-715-3084

Manas Chattopadhyay, 901 F Street NW Suite 400, Washington, DC 20004, manas_chattopadhyay@gallup.com, 202-715-3179

Anna Cahn, 901 F Street NW Suite 400, Washington, DC 20004, anna_chan@gallup.com, 402-818-1868

**APPENDIX A-QUESTIONNAIRE WORDING**
The following includes complete question wording from the Gallup Daily Tracking survey for all relevant questions included in this analysis. All parenthetical response categories were not read to the respondent but were accepted after one probe by the interviewer.

Is religion an important part of your daily life?

1. Yes
2. No
3. (DK)
4. (Refused)

In politics, as of today, do you consider yourself [(If code 1, read, A Republican, A Democrat)/(If code 2, read, a Democrat, a Republican)], or an Independent?

1. Democrat
2. Republican
3. Independent
4. (DK)
5. (Refused)

Have you, or has anyone currently living in your household, ever served in the U.S. military?
Do you approve or disapprove of the way [name of the president] is handling his job as president?

1. Approve
2. Disapprove
3. (DK)
4. (Refused)

**Employment status**

**WP10200.** Thinking about your WORK SITUATION over the past 7 days, have you been employed by an employer—even minimally like for an hour or more—from whom you receive money or goods? (This could be for one or more employers.)

(Interviewer: PAYCHECKS ARE CONSIDERED “MONEY OR GOODS”)

(Interviewer: If the respondent has trouble deciding if they are employed for an organization, probe by asking:) Do you receive money or a paycheck from them?

(Interviewer: If the respondent says they are "self-employed," code as NO, and ask next question about self-employment.)

1. Yes
2. No
3. (DK)
4. (Refused)

*Skip: (If code 1 in WP10200, Continue; Otherwise, Skip to WP10202)*

**WP10215.** In a typical week (7 days), how many hours do you work for an employer.

(Programmer: Open ended and code)

(Interviewer: This could be for one or more employers.)

(Interviewer: If respondent says "Don’t know" or has difficulty determining the number of hours worked in the last 7 days, you can help him/her calculate the 'approximate' number of hours TYPICALLY worked each week.)

(Interviewer: For example, you could ask him/her:) About how many days do you work in a typical week? AND about how many hours do you work in a
typical day? (From this, you can approximate the number of hours typically worked in a week.)

(Interviewer: If the respondent is on leave from their job and not working normal hours, re-ask the question stressing the word typical.)

1. 30 hours per week or more
2. 15 to 29 hours per week
3. 5 to 14 hours per week
4. Less than 5 hours per week
5. (DK)
6. (Refused)

Skip: (If code 1 in WP10215, Skip to WP102CK2; Otherwise, Continue)

WP10202. Again thinking about the last 7 days, were you self-employed, even minimally like for an hour or more? This means working for yourself, freelancing, or doing contract work, OR working for your own or your family’s business?

Self-employment also includes fishing, doing farm work, or raising livestock for either your own or your family’s farm or ranch.

(Interviewer: If the respondent says YES after you read the first sentence, (again, thinking about the last 7 days, were you self-employed, even minimally like for an hour or more) OR has already told you they are self-employed, you may record as YES and you do not need to read the rest of the question. If the respondent does not volunteer a "yes" while reading the first part of the question or says NO during the first sentence you MUST read the remainder of the question.)

1. Yes
2. No
3. (DK)
4. (Refused)

Skip: (If code 2-4 in WP10200 AND code 2-4 in WP10202, Skip to WP10208; If code 2-6 in WP10215 AND code 2-4 in WP10202, Skip to WP10229; Otherwise, Continue)

(Programmer: If code 1 in WP10202, ask:)

Survey Practice
**WP10216.** In a typical week (7 days), how many hours do you work as a self-employed individual?

(Programmer: Open ended and code)

(Interviewer: If respondent says "Don't know" or has difficulty determining the number of hours worked in the last 7 days, you can help him/her calculate the 'approximate' number of hours TYPICALLY worked each week.)

(Interviewer: For example, you could ask him/her:) About how many days do you work in a typical week? AND about how many hours do you work in a typical day? (From this, you can approximate the number of hours typically worked in a week.)

(Interviewer: If the respondent is on leave from their job and not working normal hours, re-ask the question stressing the word typical.)**

1. 30 hours per week or more
2. 15 to 29 hours per week
3. 5 to 14 hours per week
4. Less than 5 hours per week
5. (DK)
6. (Refused)

*Skip: (If code 2-6 in WP10216, Continue; Otherwise, Skip to Note before WP10208)*

(Programmer: If [code 2-6 in WP10216] OR [code 2-6 in WP10215 AND code 2-4 in WP10202], ask:)

**WP10229**

Do you want to work 30 hours or more per week?

1. Yes
2. No
3. (DK)
4. (Refused)

*Skip: (If code 2, 3, or 4 in WP10200 AND WP10202, Continue; Otherwise, Skip to WP102CK2)*

**WP10208.** In the past four weeks, have you actively been looking for
employment? "Actively looking" means applying for jobs, searching for jobs, and the like.

1. Yes
2. No
3. (DK)
4. (Refused)

WP10209. Would you have been able to begin work had you been offered a job within the last four weeks?

1. Yes
2. No
3. (DK)
4. (Refused)

For details about how Gallup calculates employment status categories, see http://news.gallup.com/poll/128192/gallup-employment-tracking-work.aspx
REFERENCES

Benford, R., L. Piekarksi, J. Lien, and T. Thompson. 2012. “Geographic Differences between RDD Cell and Landline Frames and Self-Report.” In The American Association for Public Opinion Research Conference. Orlando, FL.

Christian, L., M. Dimock, and S. Keeter. 2009. “Accurately Locating Where Wireless Respondents Live Requires More than a Phone Number.” Pew Research Center for People and the Press. Posted July 9th. http://pewresearch.org/pubs/1278/cell-phones-geographic-sampling-problems.

Dost, M., and K. McGeeney. 2016. “Moving without Changing Your Cellphone Number: A Predicament for Pollsters.” Pew Research Center for People and the Press. Posted August 1st. http://www.pewresearch.org/2016/08/01/moving-without-changing-your-cellphone-number-a-predicament-for-pollsters/.

Dutwin, D. 2014. “Billing ZIP Codes in Cellular Telephone Sampling.” Survey Practice 7 (4).

Kennedy, C. 2007. “Evaluating the Effects of Screening for Telephone Service in Dual Frame RDD Surveys.” Public Opinion Quarterly 71 (5): 750–71.

Montgomery, R., X. Tao, K. Wolter, S. Greby, and E. Kennedy. 2011. “Accuracy of Geographic Stratification in a Cell-Phone Survey.” In The American Association for Public Opinion Research Conference. Phoenix, AZ.

Skalland, B., and M. Khare. 2013. “Geographic Inaccuracy of Cell Phone Samples and the Effect on Telephone Survey Bias, Variance, and Cost.” Journal of Survey Statistics and Methodology 1 (1): 45–65.