National Health Statistics Reports

Number 14 ■ March 11, 2009

Wireless Substitution: State-level Estimates From the National Health Interview Survey, January–December 2007

by Stephen J. Blumberg, Ph.D., and Julian V. Luke, Division of Health Interview Statistics, National Center for Health Statistics, and Gestur Davidson, Ph.D.; Michael E. Davern, Ph.D.; Tzy-Chyi Yu, M.H.A., Ph.D.; and Karen Soderberg, M.S., the State Health Access Data Assistance Center, University of Minnesota

Abstract

Objectives—This report presents state-level estimates of the percentage of households that do not have a landline telephone, but do have at least one wireless telephone. These wireless-only households made up 14.7% of U.S. households in 2007. The report also presents state-level estimates of the percentage of adults living in wireless-only households. These wireless-only adults made up 13.6% of U.S. adults in 2007.

Methods—A two-sample modeling strategy was used to estimate the prevalence of wireless-only households and adults by state. This modeling was based on data from the 2007 National Health Interview Survey and the 2008 Current Population Survey's Annual and Social Economic Supplement.

Results—The results show that the prevalence of wireless-only households and adults in 2007 varied substantially across states. State-level estimates ranged from 5.1% (Vermont) to 26.2% (Oklahoma) of households and from 4.0% (Delaware) to 25.1% (Oklahoma) of adults. In addition, approximately one out of four adults (25.4%) living in the District of Columbia were wireless-only.

Keywords: wireless substitution • cell phones • telephone surveys • noncoverage

Introduction

The National Health Interview Survey (NHIS) is the most widely cited source for data on the number of American homes that only have wireless telephones. Every 6 months, the National Center for Health Statistics (NCHS) releases a report with the most up-to-date estimates available from the federal government concerning the size and characteristics of the wireless-only population (1). That report, published as part of the NHIS Early Release Program, presents national and regional estimates. For example, the latest results show that more than one out of every six American homes (17.5%) had only wireless telephones during the first half of 2008 (1).

Most major survey research organizations in the United States, including NCHS, have not traditionally included wireless telephone numbers when conducting random-digit-dial telephone surveys. The exclusion of households with only wireless telephones has potential implications for results from health surveys, political polls, and other research conducted using random-digit-dial methods. Indeed, the potential for bias due to incomplete coverage of the U.S. household population (that is, due to noncoverage of wireless-only and phoneless households) remains a real and growing threat to health surveys conducted only on landline telephones (2–4).

For this reason, survey systems that have relied on random-digit-dial surveys for years have been testing methods for including samples of wireless-only households. These systems include several conducted by the Centers for Disease Control and Prevention (CDC), including the Behavioral Risk Factor Surveillance System, the National Immunization Survey, and the State and Local Area Integrated Telephone Survey. These three systems collect data and produce results at the state level. For them to effectively combine samples of wireless-only households with samples of landline households from random-digit-dial surveys, state-level estimates of the prevalence of wireless-only households are needed. Yet, direct state-level estimates of this prevalence have not been available from NHIS data

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics

because the sample size of NHIS is insufficient for direct reliable annual estimates for most states.

This report presents results of modeled estimates of the prevalence of wireless-only households and wireless-only adults at the state level, using data from the 2007 NHIS and the 2008 Current Population Survey's (CPS) Annual and Social Economic Supplement (ASEC). In contrast to the NHIS, the CPS has sufficient sample size for direct reliable annual demographic estimates for all states, but does not include questions necessary to identify wireless-only households. By incorporating data from both surveys, the modeled estimates presented here take advantage of the unique strengths of both surveys. To our knowledge, these estimates are the first state-level estimates of the size of this population available from the federal government.

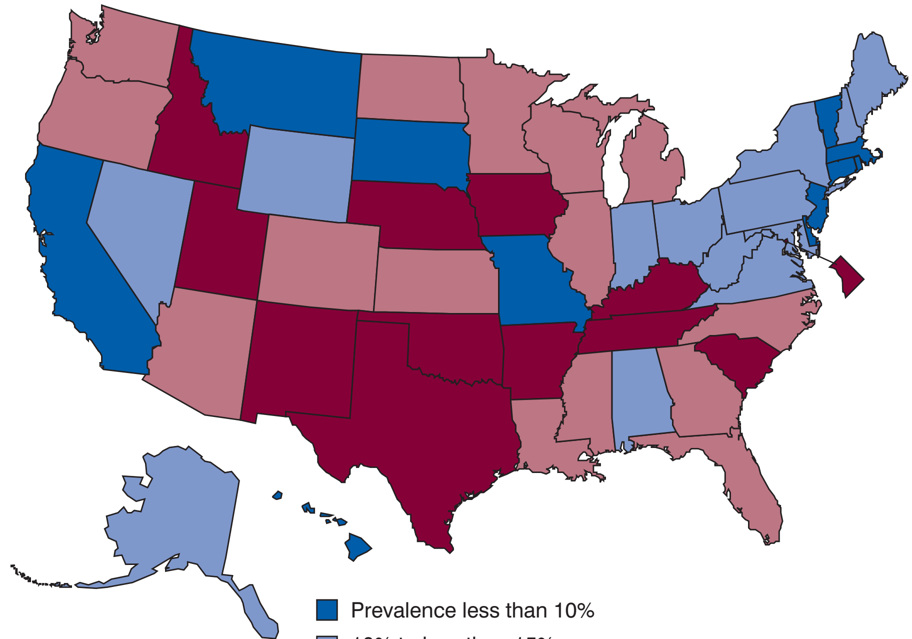

Figure 1. State-level comparisons of the percentage of wireless-only households, modeled estimates: United States, 2007

Methods

A two-sample modeling strategy was used to estimate the prevalence of wireless-only households and of adults living in such households, by state. This strategy used an optimal blend of direct estimates and synthetic estimates (5,6). First, NHIS data were used to fit multivariate regression models predicting wireless-only status using covariates from NHIS that were also available from the CPS. The model was then used with CPS data to obtain average state-level synthetic estimates (or predictions). Standard errors (SEs) for these mean state-level synthetic estimates were also obtained using Stata version 10 with the Delta method (7). Next, NHIS data were used to directly obtain state-level prevalence estimates and their corresponding (and often large) SEs. Finally, a blended overall estimate was calculated as the weighted sum of the synthetic and direct estimates for each state, where the weights reflected the relative precision of each estimate.

More detail regarding this estimation methodology is available in the "Technical Notes."

Results

Results from the two-sample modeling strategy show great variation in the prevalence of wireless-only households across states (see Figures 1 and 2). Household-level estimates ranged from a low of 5.1% in Vermont to a high of 26.2% in Oklahoma (see Table).

Other states with a high prevalence of wireless-only households include Utah (25.5%), Nebraska (23.2%), Arkansas (22.6%), Idaho (22.1%), and Iowa (22.2%). Other states with a low prevalence of wireless-only households include Connecticut (5.6%), Delaware (5.7%), South Dakota (6.4%), Rhode Island (7.9%), New Jersey (8.0%), and Hawaii (8.0%).

Similarly, results show great variation in the prevalence of wireless-only adults across states, ranging from a low of 4.0% in Delaware to a high of 25.1% in Oklahoma (see Table). An ostensibly but not significantly higher prevalence rate was observed for adults living in the District of Columbia (25.4%). Other states with a high prevalence of wireless-only adults include Utah (23.9%), Nebraska (22.4%), Kentucky (21.6%), Idaho (21.3%), and Arkansas (21.2%). Other states with a low prevalence of wireless-only adults include Vermont (4.6%), Connecticut (4.8%), Rhode Island (5.3%), Montana (5.4%), and New Jersey (6.1%).

Conclusion

Because of the absence of state-level prevalence estimates for the wireless-only population, survey researchers interested in combining state-level samples of wireless-only households with samples of landline households have relied on national or regional estimates of the relative sizes of these two populations (8). Similarly, telecommunications companies seeking greater understanding of conditions in state and local markets have relied on regional estimates of the prevalence of wireless-only households (9). The results in this report clearly show that, for many states, national and regional estimates are not sufficiently accurate for these purposes.

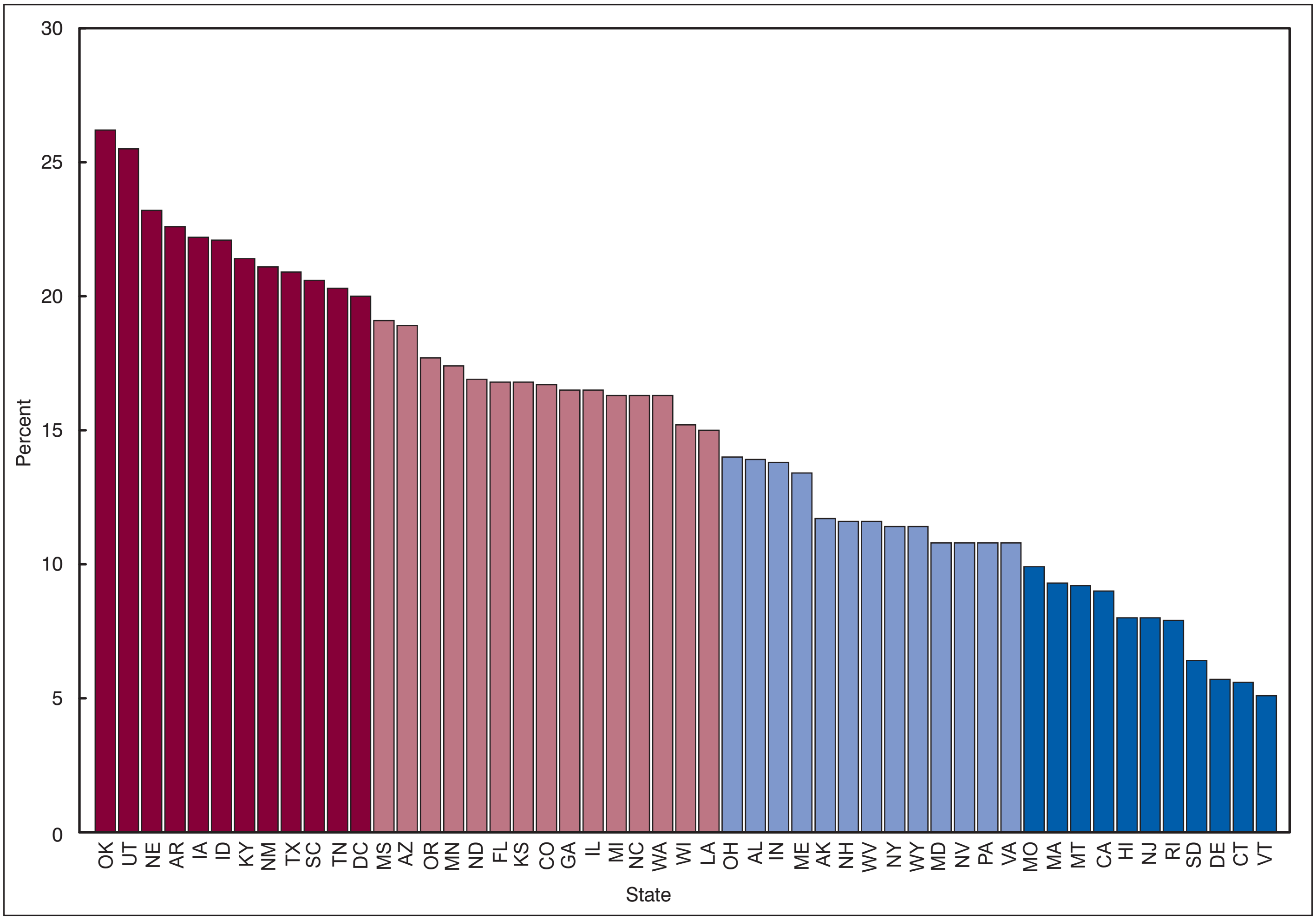

Figure 2. Modeled state-level estimates of the percentage of wireless-only households: United States, 2007

Results from the two-sample modeling strategy show great state-level variation in the prevalence of wireless-only households, even within regions. The range of prevalence exceeded 8 percentage points in the Northeast region and 20 percentage points in the South region. In fact, in the Midwest region, the state with the lowest prevalence (South Dakota, 6.4%) borders the state with the highest prevalence (Nebraska, 23.2%). Similar ranges within regions were observed for estimates of the prevalence of wireless-only adults.

Of course, for survey researchers and telecommunications companies interested in local areas, these state-level prevalence estimates still may not be sufficiently specific. For example, national estimates suggest that adults living in metropolitan areas are more likely to live in wireless-only households than are adults living in nonmetropolitan areas. Variation across local areas within a state should be expected, just as there was variation across states within a region. NCHS intends to continue working with the University of Minnesota to use two-sample modeling strategies like this one to produce estimates of telephone status for large metropolitan areas. Meanwhile, researchers may find the state-level model specifications (in the ''Technical Notes'') useful for creating completely synthetic predictions for local areas or other subpopulations of interest.

Survey researchers and telecommunication companies using the estimates presented in this report should be aware that these estimates are based on 2007 data. The number of American homes with only wireless telephones continues to grow (1). The estimate from the first half of 2008, that 17.5% of households were wireless-only, is nearly 3 percentage points higher than the estimate for the 2007 calendar year (14.7%). Similarly, the estimated prevalence of wireless-only adults has grown from 13.6% in 2007 to 16.1% in the first half of 2008. We do not know if the rates of growth in each state are comparable to the national rates, or whether they vary substantially (as did the overall prevalence rates by state). Regardless, it is very likely that the current state-level prevalence rates of wireless-only households and adults are greater than the estimates presented here.

For More Information

For more information about the implications of wireless-only households for health surveys based on landline telephone interviews, see other reports (1–4). For more information about the design, content, and use of the NHIS, please visit the NCHS website (http://www.cdc.gov/nchs/nhis.htm).

References

1. Blumberg SJ, Luke JV. Wireless substitution: Early release of estimates based on data from the National Health Interview Survey, January – June 2008. National Center for Health Statistics. 2008. Available from: http://www.cdc.gov/nchs/about/major/nhis/releases.htm.
2. Blumberg SJ, Luke JV. Coverage bias in traditional telephone surveys of low-income and young adults. Public Opinion Quarterly 71:734–49. 2007.
3. Blumberg SJ, Luke JV, Cynamon ML. Telephone coverage and health survey estimates: Evaluating the need for concern about wireless substitution. Am J Public Health 96:926–31. 2006.
4. Blumberg SJ, Luke JV, Cynamon ML, Frankel MR. Recent trends in household telephone coverage in the United States. In JM Lepkowski et al. (eds.), Advances in telephone survey methodology 56–86. New York: John Wiley and Sons, Inc. 2008.
5. Schirm AL, DiCarlo JR. Using Bayesian shrinkage methods to derive state estimates of poverty, food stamp program eligibility, and food stamp program participation. Washington, DC: Mathematica Policy Research, Inc. 1998. Available from: http://www.mathematica-mpr.com/publications/PDFs/bayesian.pdf.
6. National Research Council. Small area income and poverty estimates: Priorities for 2000 and beyond. Panel on Estimates of Poverty for Small Geographical Areas, CF Citro and G Kalton, editors. Committee on National Statistics. Washington, DC: National Academy Press. 2000.
7. Greene WH. Econometric Analysis, 5th Ed. Upper Saddle River, NJ: Prentice Hall. 2003.
8. AAPOR Cell Phone Task Force. Guidelines and considerations for survey researchers when planning and conducting RDD and other telephone surveys in the U.S. with respondents reached via cell phone numbers. American Association for Public Opinion Research. 2008. Available from: http://www.aapor.org/uploads/Final_AAPOR_Cell_Phone_TF_report_041208.pdf.
9. Petitions of Qwest Corporation for Forbearance Pursuant to 47 U.S.C. § 160(c) in the Denver, Minneapolis-St. Paul, Phoenix, and Seattle Metropolitan Statistical Areas, WC Docket No. 07–97, Memorandum Opinion and Order, FCC 08–174. Federal Communications Commission. July 25, 2008. Available from: http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-08–174A1.pdf.
10. Davern M, Jones A, Lepkowski J, Davidson G, Blewitt LA. Estimating regression standard errors with data from the Current Population Survey's public use file. Inquiry 44:211–24. 2007.

Table. Modeled state-level estimates of the percentage of wireless-only households and the percentage of adults living in wireless-only households: United States, 2007

State	Households		Adults	
	Percent	Widest plausible interval	Percent	Widest plausible interval
Alabama	13.9	9.7–18.1	12.2	8.1–16.4
Alaska	11.7	8.9–14.8	13.3	7.3–19.5
Arizona	18.9	14.5–23.1	17.1	13.6–20.4
Arkansas	22.6	18.7–26.4	21.2	16.8–25.6
California	9.0	8.1–9.8	8.4	7.7–9.1
Colorado	16.7	13.2–20.3	15.2	12.5–18.2
Connecticut	5.6	3.4–7.8	4.8	2.7–6.9
Delaware	5.7	4.8–6.8	4.0	2.8–5.4
District of Columbia	20.0	15.5–24.5	25.4	15.2–34.1
Florida	16.8	13.9–19.4	15.5	12.8–17.8
Georgia	16.5	12.9–19.9	15.0	11.6–18.1
Hawaii	8.0	6.5–9.6	8.2	7.4–8.8
Idaho	22.1	18.9–25.3	21.3	19.0–23.9
Illinois	16.5	14.1–18.7	15.2	12.8–17.1
Indiana	13.8	10.3–16.9	13.0	8.9–16.8
Iowa	22.2	9.8–34.1	18.9	7.8–29.3
Kansas	16.8	12.8–20.6	15.2	11.9–18.1
Kentucky	21.4	11.7–30.4	21.6	11.5–30.8
Louisiana	15.0	10.2–19.6	13.8	9.6–17.9
Maine	13.4	10.5–16.5	12.0	10.6–13.9
Maryland	10.8	9.1–12.6	9.8	8.3–11.5
Massachusetts	9.3	7.9–10.7	8.4	7.1–9.8
Michigan	16.3	12.7–19.7	15.3	11.6–18.7
Minnesota	17.4	14.4–20.3	16.5	14.7–18.2
Mississippi	19.1	11.4–26.3	20.3	12.6–27.0
Missouri	9.9	6.8–12.9	8.4	6.2–10.6
Montana	9.2	8.0–10.6	5.4	4.5–6.4
Nebraska	23.2	13.2–32.7	22.4	12.7–31.2
Nevada	10.8	8.8–13.0	10.1	9.0–11.3
New Hampshire	11.6	9.2–14.3	8.9	7.2–11.0
New Jersey	8.0	6.0–10.0	6.1	4.8–7.5
New Mexico	21.1	11.3–29.6	20.5	10.4–28.8
New York	11.4	10.0–13.0	10.6	9.4–12.2
North Carolina	16.3	13.6–19.0	14.8	12.3–17.3
North Dakota	16.9	6.7–27.2	18.1	4.4–32.2
Ohio	14.0	11.3–16.6	13.1	11.0–15.3
Oklahoma	26.2	12.9–38.8	25.1	14.6–34.6
Oregon	17.7	14.5–20.8	18.1	15.0–20.8
Pennsylvania	10.8	8.6–13.0	9.2	7.3–11.2
Rhode Island	7.9	0.1–15.6	5.3	0.3–11.0
South Carolina	20.6	14.5–26.0	19.2	13.8–24.0
South Dakota	6.4	5.7–7.1	6.8	6.1–7.6
Tennessee	20.3	16.1–23.4	20.8	14.9–25.2
Texas	20.9	18.3–23.0	19.5	17.0–21.2
Utah	25.5	16.9–32.8	23.9	15.2–30.9
Vermont	5.1	4.9–5.4	4.6	4.5–4.9
Virginia	10.8	8.8–12.9	10.0	7.9–12.2
Washington	16.3	12.4–20.2	15.6	12.2–19.0
West Virginia	11.6	8.3–14.5	10.6	4.6–16.1
Wisconsin	15.2	11.9–18.4	13.6	10.8–16.3
Wyoming	11.4	10.8–12.2	13.0	12.3–14.2

DATA SOURCES: CDC/NCHS, National Health Interview Survey, 2007, and U.S. Census Bureau, Current Population Survey, Annual and Social Economic Supplement, 2008. Estimates were calculated by the State Health Access Data Assistance Center, University of Minnesota.

NOTE: Please refer to the ''Technical Notes'' for a description of the calculation of the ''widest plausible interval.''

Technical Notes

Data sources

The state-level estimates presented in this report are based on data from the 2007 National Health Interview Survey (NHIS) and the 2008 CPS Annual and Social Economic Supplement. NHIS is a multipurpose health survey conducted by CDC's NCHS. The CPS is an annual demographic survey conducted by the U.S. Census Bureau for the Bureau of Labor Statistics.

NHIS is an annual multistage probability household survey of a large sample of households drawn from the civilian noninstitutionalized household population of the United States. This face-to-face survey interview is administered by trained field representatives from the U.S. Census Bureau. NHIS interviews are conducted continuously throughout the year to collect information on health status, health-related behaviors, and health care utilization. The survey also includes information about household telephones and whether anyone in the household has a wireless telephone (also known as a cellular telephone, cell phone, or mobile phone).

The sample for NHIS is stratified by state, which allows use of NHIS for producing state-level estimates. However, the current NHIS sample size is not sufficient to provide reliable annual state-level estimates for most states. In 2007, household telephone status information was obtained for 29,079 households; of these, 28,492 had sufficient nonmissing data for the covariates to be included in the multivariate regression analyses presented here.

The CPS is a multistage probability household survey that provides data on labor force participation and unemployment. Data are collected through a combination of face-to-face and telephone interviews. The ASEC is added one time per year to the monthly CPS (from February through April) and is used to produce household income, family poverty, and health insurance coverage estimates. The reference period for these ASEC items is the prior calendar year; the 2008 CPS ASEC uses 2007 as its reference period. The CPS ASEC is both nationally and state representative and has included approximately 78,000 households per year since 2000.

NHIS and CPS sampling weights adjust for the probability of selection of each household, and they are adjusted for nonresponse. The results reported in this report are based on weighted estimates. StataSE v.10 software was used to account for the complex survey designs.

Definition of a wireless-only household

For each family contacted by NHIS, one adult family member was asked whether ''you or anyone in your family has a working cellular telephone.'' A family can be an individual or a group of two or more related persons living together in the same housing unit. Thus, a family can consist of only one person, and more than one family can live in a household (including, for example, a household where there are multiple single-person families, as when unrelated roommates are living together).

To produce the statistics for this report, families were identified as wireless families if anyone in the family had a working cellular telephone. Households were identified as wireless-only if they included at least one wireless family and if there were no working landline telephones inside the household. To determine if there was a working landline telephone inside the household, survey respondents were asked if there was ''at least one phone inside your home that is currently working and is not a cell phone.''

Household telephone status (rather than family telephone status) is used in this report because most telephone surveys draw samples of households rather than families. Adults are identified as wireless-only if they live in a wireless-only household. Individual ownership or use of cellular telephones is not determined.

Two-sample state-level estimation

The goal was to develop a robust set of blended direct and synthetic estimates of the wireless-only population. We used both the NHIS and the CPS Annual and Social Economic Supplement in a two-sample modeling approach (5,6). Specifically, we first fitted a multivariate regression model using NHIS data. NHIS is the only survey that provides information on wireless-only households. We then used data from the CPS to derive state-level predictions through the NHIS-fitted model. The CPS has a larger sample size and a survey design that produces reliable, state-representative estimates. We then blended the direct estimates from the NHIS and the modeled estimates from the CPS data and NHIS model to yield a set of improved state-level estimates. We undertook this modeling exercise separately for two units of analysis: for the household and for adults aged 18 years and older. For ease of exposition, we describe the process undertaken using the household as the unit of analysis. The process was the same when adults were the unit of analysis.

Formally, this two-sample strategy has five steps. First, using the multinomial logistic (MNL) regression command for survey data in Stata 10.1 (''svy: mlogit'') and confidential internal NHIS data files, we fitted a fixed-effects model (at the level of state) on this national sample of 28,492 households. The three categories for our multinomial dependent variable were wireless-only, landline (with or without the presence of a wireless telephone), and phoneless. With access to the detailed confidential NHIS sample design and using Stata's survey suite commands, we were able to take full account of the complex survey design to obtain robust standard errors for this multivariate model.

Second, using publicly available CPS data files, we recycled the CPS data on all our model covariates (i.e., recycled predictions) through our NHIS-fitted model to obtain the average state model-based estimates. These average state model-based estimates were obtained as the sum of the model-estimated probabilities of each household in a state being wireless-only divided by the number of sample observations at the level of the state. Symbolically denoting the j^{th} state's estimated rate of wireless-only (*wo*) using the MNL model as $\hat{P}_{wo,j}^{cps}$, we refer to these $\hat{P}_{wo,j}^{cps}$, as our *synthetic estimates*.

Third, we used the Stata command ''adjust'' to obtain the standard errors of these mean state-level model predictions (SE_p^{cps}). With access to the publicly available CPS data files only, we could not take full account of the complex survey design when calculating these standard errors. Instead, we used the Delta method (7) with the Stata command ''adjust,'' and we identified the lowest level of identifiable geography in the publicly available CPS data files as the strata variable. This alternative method has been shown to yield standard errors for multivariate regression models of dichotomous dependent variables that are very close to those obtained when the full confidential survey design is available (10). As such, these standard errors are likely to fully reflect both sampling and model-based imputation errors.

Fourth, we used the NHIS survey data and Stata (''svy: mean'') to obtain the direct prevalence estimate for the wireless-only variable for each state, denoted by $DE_{wo,j}$, along with the standard error of these means, $SE_{p,j}^{de}$. We call these values of $DE_{wo,j}$ our *direct estimates* (Table I). As noted earlier, we were able to take full account of the complex NHIS design to obtain these standard errors.

Finally, we formed the blended wireless-only estimate as the weighted sum of the synthetic estimate and direct estimate for each state, where the weights reflect the relative precision of each state's pair of synthetic and direct estimates, as:

$$\hat{P}_{\text{wo, j}}^{\text{blended}} = \hat{P}_{\text{wo, j}}^{\text{cps}} \left[\frac{1/(SE_{p,j}^{cps})^2}{1/(SE_{p,j}^{cps})^2 + 1/(SE_{p,j}^{de})^2} \right] +$$

$$DE_{wo,j}^{de} \left[\frac{1/(SE_{p,j}^{de})^2}{1/(SE_{p}^{cps})^2 + 1/(SE_{p,j}^{de})^2} \right].$$

We used a MNL regression model with the three categories of wireless-only, landline, and phoneless rather than a binomial logistic regression model with the two categories of wireless-only and ''all others.'' We used three categories because, though the prevalence of phoneless households is quite small, our MNL regression analyses revealed that the phoneless equation's coefficients were almost as large and significant as those for the wireless-only equation. Thus, combining the landline and phoneless categories into a single ''all others'' category and estimating a binomial logistic regression for wireless-only would have resulted in heterogeneity within this combined ''all others'' category. As a consequence, the wireless-only equation's coefficients would likely have been biased.

We tested the appropriateness of our use of a MNL regression model rather than the use of much more computationally demanding multinomial probit regression models (MNP). As part of their estimation approach, MNL models make the assumption—referred to as Independence from Irrelevant Alternatives (IIA)—that the coefficients in a MNL model will not depend on whether an outcome-level within the multinomial response variable is included in the estimation or excluded (and its data removed). MNP does not make this assumption and consequently it is often recommended as the appropriate estimator. MNP is very computationally demanding, however. The IIA assumption underlying MNL can be tested empirically, and there are two variants of it formulated as Hausman tests, plus a third IIA test known as the Small-Hsiao test. In our particular modeling example, the IIA tests can be simply described as assessing whether the coefficients of the wireless-only equation in a MNL model are significantly different from the coefficients estimated from a wireless-only binomial logistic regression model that excludes the phoneless data. Two of the three tests favored the IIA assumption. Therefore, we believe our use of a MNL model is appropriate.

Selection of a fixed-effects model

We note that the NHIS has a relatively large sample size—although not designed for representative annual state-level estimates—and that the CPS has an even greater sample size and is designed to produce representative annual state-level estimates. Additionally, we note that, with our model specification, we were able to account for significant variation in wireless-only households. Given all these considerations, we believe it was appropriate that differences in each state's direct and synthetic estimates should reflect *only* the differences across NHIS and CPS surveys in their covariate means, $\overline{X}$, where these covariates have been shown empirically to be significant predictors of whether households are wireless-only. For this reason, we believe that a fixed effect modeling strategy is a better specification than a nonfixed effect strategy would have been.

Assume that after estimating our fixed effects model with NHIS data we had generated a mean predicted NHIS value for each state by recycling NHIS data through the model. Let these values be denoted as $\beta_{NHIS}^{recycled}$. We in fact generated a mean predicted value for each state by recycling the CPS data through the model. Let these values be denoted as $\beta_{CPS}^{recycled}$. Finally, we also used NHIS data to produce a direct prevalence estimate for each state. Let these values be denoted as DE_{NHIS}.

Given these three estimates, for any state, we can form the ratio:

$$\frac{[DE_{NHIS} - \beta_{NHIS}^{recycled}] + [\beta_{NHIS}^{recycled} - \beta_{CPS}^{recycled}]}{DE_{NHIS} - \beta_{CPS}^{recycled}},$$

which is identically equal to 1.0. By the first order, conditions of the logistic

Table I. Direct state-level estimates of the percentage of wireless-only households and the percentage of adults living in wireless-only households: United States, 2007

State	Households		Adults	
	Percent	95% confidence interval	Percent	95% confidence interval
Alabama	12.5	5.8–19.2	11.7	5.0–18.4
Alaska	9.8	1.1–18.6	11.2	1.8–20.6
Arizona	18.8	13.4–24.2	17.1	13.0–21.1
Arkansas	22.3	17.5–27.0	20.7	14.8–26.5
California	8.4	7.3–9.4	7.7	6.8–8.6
Colorado	18.1	12.2–24.1	17.0	10.9–23.2
Connecticut	6.1	2.9–9.2	5.0	1.9–8.1
Delaware	7.7	0.3–15.0	5.0	0.0–10.4
District of Columbia	23.0	15.3–30.8	25.7	17.1–34.3
Florida	15.8	13.1–18.5	14.5	12.0–17.0
Georgia	15.5	11.6–19.4	13.7	10.0–17.3
Hawaii	7.6	5.2–10.0	8.2	7.4–9.0
Idaho	23.2	18.5–28.0	21.5	17.6–25.4
Illinois	16.3	13.7–18.9	15.0	12.9–17.2
Indiana	13.8	10.5–17.2	13.5	9.2–17.8
Iowa	23.6	10.5–36.7	21.4	9.9–33.0
Kansas	15.8	11.7–20.0	13.9	10.6–17.2
Kentucky	22.0	13.1–30.9	22.8	13.2–32.3
Louisiana	13.8	8.3–19.2	13.2	8.4–17.9
Maine	17.3	4.7–29.8	15.7	4.0–27.3
Maryland	10.1	7.3–12.9	8.9	6.4–11.5
Massachusetts	8.9	6.9–10.9	8.0	6.0–10.0
Michigan	17.4	13.0–21.8	16.5	11.9–21.2
Minnesota	16.2	12.3–20.0	15.7	13.5–17.8
Mississippi	18.2	10.9–25.5	19.1	12.3–25.8
Missouri	9.9	6.2–13.7	8.3	5.4–11.3
Montana	5.0	0.5–9.4	3.1	1.8–4.3
Nebraska	22.3	12.0–32.6	20.8	12.2–29.5
Nevada	10.3	5.8–14.7	9.0	7.4–10.5
New Hampshire	12.1	0.0–25.3	9.1	0.0–19.8
New Jersey	7.9	5.3–10.5	5.9	4.1–7.6
New Mexico	20.3	12.6–27.9	19.4	11.7–27.1
New York	12.8	9.2–16.5	12.1	8.5–15.8
North Carolina	15.9	12.2–19.6	14.6	11.2–18.1
North Dakota	14.3	0.0–30.4	15.3	0.0–32.0
Ohio	14.5	10.8–18.2	13.7	10.3–17.1
Oklahoma	27.4	14.2–40.6	26.5	16.3–36.7
Oregon	17.3	13.5–21.1	17.4	14.3–20.5
Pennsylvania	11.0	8.0–14.0	9.6	6.7–12.5
Rhode Island	6.8	0.0–14.4	4.4	0.0–10.5
South Carolina	19.6	14.2–24.9	18.3	13.2–23.4
South Dakota	4.8	3.0–6.6	5.5	4.3–6.6
Tennessee	19.9	16.9–23.0	20.2	15.9–24.5
Texas	20.6	18.3–22.8	18.7	16.8–20.7
Utah	24.0	17.3–30.7	22.2	15.6–28.8
Vermont	3.5	0.0–8.8	2.8	0.0–6.3
Virginia	11.9	8.3–15.6	10.9	7.0–14.7
Washington	16.6	11.2–22.1	16.1	11.4–20.7
West Virginia	10.0	7.2–12.9	9.6	4.6–14.7
Wisconsin	16.3	12.1–20.4	15.2	11.3–19.2
Wyoming	14.2	4.6–23.7	16.8	10.0–23.5

0.0 Quantity more than zero but less than 0.05.

DATA SOURCE: CDC/NCHS, National Health Interview Survey, 2007.

model (for binomial or MNL regressions), when a model is a fixed effects model at the state level, then for each state, $DE_{NHIS} = \beta^{recycled}_{NHIS}$, in which case, the above equation becomes:

$$\frac{\beta^{recycled}_{NHIS} - \beta^{recycled}_{CPS}}{DE_{NHIS} - \beta^{recycled}_{CPS}} \equiv 1.0.$$

In other words, in our state-level fixed effects models, the difference of any state's direct and synthetic estimates in the estimated proportion of households that are wireless-only reflects—solely—the weighted differences across the NHIS and CPS surveys in their covariate means, $\overline{X}$. That is, differences in direct and synthetic estimates are a weighted average across NHIS and CPS surveys of their covariate means, $\overline{X}$, with these weights being the size of the coefficients in our MNL regression model.

Selection of variables used in the models

The wireless-only classification is a *collective* attribute or characteristic of a household. It is either present in a household or not present, and its significance is that no member of that household can be contacted with a landline associated with that household. (By contrast, all persons in a landline household can be contacted with a landline, in the telephone-survey-based sense that any one household member contacted on a landline can tell the survey interviewer about himself or herself and all others in the household.) As a collective household characteristic, the wireless-only status of households is best modeled using variables that are also measured at the household level. We also used household-level variables when we estimated our adult-level models. That is, in the adult-level models, we looked for characteristics of the household that predicted whether that adult resides within a wireless-only household.

Given the nature of our two-sample modeling approach, the variables that could be used in the model had to be available in both the NHIS and CPS surveys. In addition, we required identical coding of all responses in these two sets of variables. We used previous work on predictors of telephone status (4) as a starting point for choosing our variables. In addition, as suggested by their work, we tested the importance and significance of several interactions between variables for home ownership, age ranges within the household, and number of persons in the household.

In addition, we expected that there would be a direct relationship between the prevalence of wireless-only households in a state and the number of wireless telephones per capita. From the Federal Communications Commission's Automated Reporting Management Information System database, we obtained the number of wireless telephone subscriptions as of December 2006 and June 2007, and we divided these respectively by the U.S. Census Bureau's July 1, 2006, and July 1, 2007, population estimates. We then entered these two state-wide values of wireless telephones per capita into the NHIS data set in conjunction with the NHIS variable that designates whether a household participated in the survey during the first half of the year or the second half, to form one overall state-specific cell subscriber variable. This state-level variable could be used in a fixed-effects model because there was variation in it within a state, and due to rapidly rising numbers of wireless telephone subscriptions, there was adequate variability. Indeed, the effect of this ''within-state'' estimator—it relates variation in a state-level variable within the individual states to variation in wireless-only prevalence within the individual states—was quantitatively large, positive, and significant. For the household-level model, $\beta = 5.05$ and $p = .015$.

MNL regression results

Table II presents the MNL regression results for the final household-level model for the wireless-only equation. Table III presents the MNL regression results for the final adult-level model for the wireless-only equation.

Variability of the estimates

For policy purposes, it's natural to seek some quantitative measure of the level of uncertainty surrounding the estimates, analogous to a 95% confidence interval (CI). Of course, these blended estimates are a weighted average of two random variables, where the weights are complex functions of each of the two estimates' SEs and thus these weights must be considered themselves to be random variables. This high level of complexity holds even if we assume, appropriately, that the NHIS and CPS survey samples are independent.

A full bootstrap procedure involving the simultaneous re-sampling of both the estimation sample (NHIS) *and* the prediction sample (CPS) could be conceptualized as a complex way of obtaining 95% CI for our blended estimates. Given this high level of complexity, we instead derived a simple quantitative measure of the level of uncertainty about these blended estimates, which due to its construction we refer to as the *widest plausible interval* for our blended estimates. Our direct estimates, $DE_{wo,j}$, have lower and upper 95% CI values, denoted as $DE^{L}_{wo,j}$, and $DE^{U}_{wo,j}$, and our synthetic estimates, $\hat{P}^{cps}_{wo,j}$, similarly have lower and upper 95% CI values, denoted $\hat{P}^{cps,L}_{wo,j}$, and $\hat{P}^{cps,U}_{wo,j}$. A maximal difference in our blended estimates could be conceived of as arising from taking the upper bounds of both our direct estimates and synthetic estimates and weighting these two upper bound values by their precision, as is done in our overall blended estimate, and then taking the lower bounds of both our direct estimates and synthetic estimates and again weighting these two lower bound values by their precision. The weights—although complex functions of SEs—are considered fixed for this purpose.

Table II. Multinomial logistic regression results for the fixed-effects household-level model

Predictor	Coefficient	Standard error	*t*-value	*p*-value
Age				
All adults in the household are less than or equal to 30 years of age	1.866711	0.089851	20.78	<.001
Household includes an adult 65 years or older	–1.539320	0.124413	–12.37	<.001
All adults in the household are between 31 and 44 years of age, inclusive	0.415542	0.073732	5.64	<.001
Home ownership status				
Renting	1.116378	0.068147	16.38	<.001
Interaction of "renting" and "all adults in the household are less than or equal to 30 years of age"	–0.565510	0.114927	–4.92	<.001
Sex				
All adults in the household are male	0.892477	0.068968	12.94	<.001
All adults in the household are female	0.319934	0.072536	4.41	<.001
Household structure				
Household includes unrelated adults	0.609820	0.078856	7.73	<.001
Household includes children under 18 years of age	–0.246420	0.087131	–2.83	0.005
Household includes only one or two adults	–0.158120	0.097213	–1.63	0.105
Race/ethnicity				
All persons in the household are Hispanic	0.406733	0.065491	6.21	<.001
All persons in the household are non-Hispanic black	–0.198420	0.074202	–2.67	0.008
Household poverty status				
Near-poor household	0.326522	0.061096	5.34	<.001
Poor household	0.241867	0.083452	2.90	0.004
Other demographics				
Household includes at least one adult with a job or business	0.402410	0.076572	5.26	<.001
At least one person in the household has a 4-year college degree or higher education	–0.174650	0.050613	–3.45	0.001
Number of persons in the household	–0.104420	0.037508	–2.78	0.006
State-level estimate of the number of wireless subscribers per capita				
Estimate from December 2006 or June 2007, depending on date of interview	5.047416	2.055141	2.46	0.015
State				
Alabama	1.272225	0.238382	5.34	<.001
Alaska	1.417395	0.469514	3.02	0.003
Arizona	1.607334	0.256893	6.26	<.001
Arkansas	2.014621	0.200072	10.07	<.001
Colorado	1.179756	0.197636	5.97	<.001
Connecticut	–0.154310	0.356241	–0.43	0.665
Delaware	–0.155140	0.173824	–0.89	0.373
District of Columbia	–2.797410	1.508089	–1.85	0.065
Florida	1.165666	0.133942	8.70	<.001
Georgia	1.232165	0.171628	7.18	<.001
Hawaii	0.252015	0.169125	1.49	0.137
Idaho	2.273645	0.333961	6.81	<.001
Illinois	1.296925	0.173017	7.50	<.001
Indiana	1.497265	0.331355	4.52	<.001
Iowa	2.035218	0.469920	4.33	<.001
Kansas	1.393183	0.218114	6.39	<.001
Kentucky	1.879856	0.413259	4.55	<.001
Louisiana	1.079136	0.231553	4.66	<.001
Maine	1.782551	0.380313	4.69	<.001
Maryland	0.486007	0.152081	3.20	0.002
Massachusetts	0.344432	0.136055	2.53	0.012
Michigan	1.674510	0.273125	6.13	<.001
Minnesota	1.563082	0.239372	6.53	<.001
Mississippi	2.179626	0.340533	6.40	<.001
Missouri	0.672953	0.304624	2.21	0.028
Montana	0.829946	0.351831	2.36	0.019
Nebraska	1.866004	0.333427	5.60	<.001
Nevada	0.379044	0.156361	2.42	0.016

Table II. Multinomial logistic regression results for the fixed-effects household-level model—Con.

Predictor	Coefficient	Standard error	*t*-value	*p*-value
New Hampshire	1.136946	0.255660	4.45	<.001
New Jersey	0.170389	0.202940	0.84	0.402
New Mexico	2.004247	0.667762	3.00	0.003
New York	0.530698	0.122055	4.35	<.001
North Carolina	1.333125	0.185659	7.18	<.001
North Dakota	1.243878	0.417006	2.98	0.003
Ohio	1.229178	0.213456	5.76	<.001
Oklahoma	2.236250	0.406260	5.50	<.001
Oregon	1.615444	0.228845	7.06	<.001
Pennsylvania	1.048467	0.266023	3.94	<.001
Rhode Island	0.564256	0.622844	0.91	0.366
South Carolina	1.929321	0.264777	7.29	<.001
South Dakota	0.228156	0.257332	0.89	0.376
Tennessee	1.119440	0.262129	4.27	<.001
Texas	1.454220	0.149066	9.76	<.001
Utah	2.551921	0.344626	7.40	<.001
Vermont	0.750223	0.479845	1.56	0.119
Virginia	0.987165	0.198595	4.97	<.001
Washington	1.195146	0.204184	5.85	<.001
West Virginia	2.273231	0.507029	4.48	<.001
Wisconsin	1.757524	0.416058	4.22	<.001
Wyoming	0.720274	0.127978	5.63	<.001
Constant				
Value	–7.645990	1.725050	–4.43	<.001

DATA SOURCES: CDC/NCHS, National Health Interview Survey (NHIS), 2007, and Federal Communications Commission, Automated Reporting Management Information System, 2006–2007.
NHIS sample size = 28,492 households.

In equation form, we have for our widest plausible interval:

$$\hat{P}^{\text{blended,L}}_{\text{wo, j}} = \hat{P}^{\text{cps,L}}_{\text{wo, j}} \left[\frac{1/(SE^{cps}_{p,j})^2}{1/(SE^{cps}_{p,j})^2 + 1/(SE^{de}_{p,j})^2} \right] + DE^{de,L}_{wo,j} \left[\frac{1/(SE^{de}_{p,j})^2}{1/(SE^{cps}_{p})^2 + 1/(SE^{de}_{p,j})^2} \right],$$

and

$$\hat{P}^{\text{blended,U}}_{\text{wo, j}} = \hat{P}^{\text{cps,U}}_{\text{wo, j}} \left[\frac{1/(SE^{cps}_{p,j})^2}{1/(SE^{cps}_{p,j})^2 + 1/(SE^{de}_{p,j})^2} \right] + DE^{de,U}_{wo,j} \left[\frac{1/(SE^{de}_{p,j})^2}{1/(SE^{cps}_{p})^2 + 1/(SE^{de}_{p,j})^2} \right].$$

Clearly, this widest plausible interval is a very conservative measure of uncertainty because it uses the upper bounds of *both* direct and synthetic estimates simultaneously and the lower bounds of *both* direct and synthetic estimates simultaneously. Although they are not CIs and almost certainly exceed in width the CIs that might be obtained from a full bootstrapping of the blended estimates, we believe these estimates of uncertainty nevertheless serve a useful purpose. These estimates are included in the Table.

Calibration of the estimates

The values of our blended estimates have been adjusted to give an overall weighted average that equals the overall NHIS wireless-only prevalence rate. This process is known in the small area estimation literature as ''raking'' the blended estimates, and it is a process usually undertaken when the state-weighted average prevalence rate from the blended methodology does not perfectly match the overall directly estimated national rate (6).

First, we calculated each state's total weight from the CPS data, formed as the product of the state's average survey-weight times the state's sample size in the CPS. Next, we multiplied each state's total weight by its blended estimate. Then, we summed these state products and divided this sum by the total of all the states' CPS weights. This yielded a rate of .1416 for households. Finally, we formed the ratio of the overall national NHIS rate to the state-weighted blended estimate rate, which is the ''raking factor.'' This ''raking factor'' was applied to each state's blended estimate to arrive at a final raked blended estimate. The direct estimate from the NHIS was .147 for households, and this yielded a ''raking factor'' of 1.0381. The raking factor for the adult estimates was 1.0599.

The small size of these raking factors (i.e., the discrepancy between the directly estimated national rate and the initial state-weighted average prevalence rate from the blended methodology) provides empirical support for the appropriateness of our modeling procedures. Bias introduced by the modeling procedures, if any, was minor.

Table III. Multinomial logistic regression results for the fixed-effects adult-level model

Predictor	Coefficient	Standard error	*t*-value	*p*-value
Age				
All adults in the household are less than or equal to 30 years of age	1.799508	0.090836	19.81	<.001
Household includes an adult 65 years or older	–1.382090	0.134494	–10.28	<.001
All adults in the household are between 31 and 44 years of age, inclusive	0.301825	0.072248	4.18	<.001
Home ownership status				
Renting	1.143872	0.067472	16.95	<.001
Interaction of "renting" and all "adults in the household are less than or equal to 30 years of age"	–0.565430	0.112602	–5.02	<.001
Sex				
All adults in the household are male	0.873861	0.073431	11.90	<.001
All adults in the household are female	0.277123	0.073664	3.76	<.001
Household structure				
Household includes unrelated adults	0.616546	0.079639	7.74	<.001
Household includes children under 18 years of age	–0.208900	0.091222	–2.29	0.023
Household includes only one or two adults	–0.155960	0.094759	–1.65	0.101
Race/ethnicity				
All persons in the household are Hispanic	0.356096	0.072708	4.90	<.001
All persons in the household are non-Hispanic black	–0.128510	0.075034	–1.71	0.088
Household poverty status				
Near-poor household	0.362912	0.063380	5.73	<.001
Poor household	0.363780	0.083803	4.34	<.001
Other demographics				
Household includes at least one adult with a job or business	0.473183	0.081272	5.82	<.001
At least one person in the household has a 4-year college degree or higher education	–0.194480	0.054364	–3.58	<.001
Number of persons in the household	–0.111140	0.037656	–2.95	0.003
State-level estimate of the number of wireless subscribers per capita				
Estimate from December 2006 or June 2007, depending on date of interview	4.280246	2.203306	1.94	0.053
State				
Alabama	1.091400	0.268269	4.07	<.001
Alaska	1.617397	0.555650	2.91	0.004
Arizona	1.480033	0.258784	5.72	<.001
Arkansas	1.918280	0.219244	8.75	<.001
Colorado	1.105028	0.189786	5.82	<.001
Connecticut	–0.250100	0.381567	–0.66	0.513
Delaware	–0.493910	0.309229	–1.60	0.111
District of Columbia	–1.915190	1.650917	–1.16	0.247
Florida	1.092798	0.130196	8.39	<.001
Georgia	1.161430	0.177353	6.55	<.001
Hawaii	0.385860	0.088463	4.36	<.001
Idaho	2.182128	0.354031	6.16	<.001
Illinois	1.250206	0.185081	6.75	<.001
Indiana	1.443714	0.373140	3.87	<.001
Iowa	1.624521	0.558052	2.91	0.004
Kansas	1.273341	0.206036	6.18	<.001
Kentucky	1.865205	0.423037	4.41	<.001
Louisiana	1.073218	0.210920	5.09	<.001
Maine	1.586943	0.389421	4.08	<.001
Maryland	0.489238	0.145614	3.36	0.001
Massachusetts	0.314262	0.141998	2.21	0.028
Michigan	1.624691	0.291254	5.58	<.001
Minnesota	1.489495	0.233542	6.38	<.001
Mississippi	2.214357	0.343132	6.45	<.001
Missouri	0.512488	0.277719	1.85	0.066
Montana	0.252518	0.388014	0.65	0.516
Nebraska	1.856366	0.359632	5.16	<.001
Nevada	0.357242	0.124442	2.87	0.004
New Hampshire	0.782579	0.273309	2.86	0.004
New Jersey	–0.049300	0.182665	–0.27	0.787
New Mexico	1.958091	0.645469	3.03	0.003
New York	0.528292	0.111597	4.73	<.001
North Carolina	1.192146	0.193867	6.15	<.001
North Dakota	1.502527	0.487668	3.08	0.002
Ohio	1.190100	0.208526	5.71	<.001

Table III. Multinomial logistic regression results for the fixed-effects adult-level model—Con.

Predictor	Coefficient	Standard error	*t*-value	*p*-value
Oklahoma	2.133292	0.383744	5.56	<.001
Oregon	1.598406	0.239406	6.68	<.001
Pennsylvania	0.874146	0.278287	3.14	0.002
Rhode Island	0.136707	0.670396	0.20	0.839
South Carolina	1.794088	0.254067	7.06	<.001
South Dakota	0.300735	0.273356	1.10	0.272
Tennessee	1.273948	0.266752	4.78	<.001
Texas	1.404436	0.149124	9.42	<.001
Utah	2.377257	0.371590	6.40	<.001
Vermont	0.576505	0.515766	1.12	0.265
Virginia	0.901419	0.225554	4.00	<.001
Washington	1.167482	0.200486	5.82	<.001
West Virginia	2.015832	0.648016	3.11	0.002
Wisconsin	1.532290	0.430925	3.56	<.001
Wyoming	0.991193	0.132367	7.49	<.001
Constant				
Value	–7.029340	1.848191	–3.80	<.001

DATA SOURCES: CDC/NCHS, National Health Interview Survey (NHIS), 2007, and Federal Communications Commission, Automated Reporting Management Information System, 2006–2007.
NHIS sample size = 53,770 adults.

Acknowledgments

This research was supported by a grant from the Robert Wood Johnson Foundation to the State Health Access Data Assistance Center at the University of Minnesota. The authors are solely responsible for the content of this report. Gestur Davidson of the University of Minnesota was the primary author of the ''Technical Notes.'' We thank Jane Gentleman, Graham Kalton, Chris Moriarity, and Allen Schirm for their comments on earlier versions of the ''Technical Notes.''

Suggested citation

Blumberg SJ, Luke JV, Davidson G, Davern ME, Yu T, Soderberg K. Wireless substitution: State-level estimates from the National Health Interview Survey, January–December 2007. National health statistics reports; no 14. Hyattsville, MD: National Center for Health Statistics. 2009.

U.S. DEPARTMENT OF
HEALTH & HUMAN SERVICES

Centers for Disease Control and Prevention
National Center for Health Statistics
3311 Toledo Road
Hyattsville, MD 20782

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

To receive this publication regularly, contact the National Center for Health Statistics by calling 1–800–232–4636
E-mail: cdcinfo@cdc.gov
Internet: http://www.cdc.gov/nchs

DHHS Publication No. (PHS) 2009–1250
CS200764-A
T33564 (03/2009)

FIRST CLASS
POSTAGE & FEES PAID
CDC/NCHS
PERMIT NO. G-284