Are the Current Population Survey Uninsurance Estimates Too High? An Examination of the Imputation Process

Michael Davern, Holly Rodin, Lynn A. Blewett, and Kathleen Thiede Call

Research Objective. To determine whether the imputation procedure used to replace missing data by the U.S. Census Bureau produces bias in the estimates of health insurance coverage in the Current Population Survey's (CPS) Annual Social and Economic Supplement (ASEC).

Data Source. 2004 CPS-ASEC.

Study Design. Eleven percent of the respondents to the monthly CPS do not take the ASEC supplement and the entire supplement for these respondents is imputed by the Census Bureau. We compare the health insurance coverage of these "full-supplement imputations" with those respondents answering the ASEC supplement. We then compare demographic characteristics of the two groups and model the likelihood of having insurance coverage given the data are imputed controlling for demographic characteristics. Finally, in order to gauge the impact of imputation on the uninsurance rate we remove the full-supplement imputations and reweight the data, and we also use the multivariate regression model to simulate what the uninsurance rate would be under the counter-factual simulation that no cases had the full-supplement imputation.

Population Studied. The noninstitutionalized U.S. population under 65 years of age in 2004.

Data Extraction Methods. The CPS-ASEC survey was extracted from the U.S. Census Bureau's FTP web page in September of 2004 (http://www.bls.census.gov/ferretftp.htm).

Principal Findings. In the 2004 CPS-ASEC, 59.3 percent of the full-supplement imputations under age 65 years had private health insurance coverage as compared with 69.1 percent of the nonfull-supplement imputations. Furthermore, full-supplement imputations have a 26.4 percent uninsurance rate while all others have an uninsurance rate of 16.6 percent. Having imputed data remains a significant predictor of health insurance coverage in multivariate models with demographic controls. Both our reweighting strategy and our counterfactual modeling show that the uninsured rate is approximately one percentage point higher than it should be for people under 65 (i.e., approximately 2.5 million more people are counted as uninsured due to this imputation bias).

Conclusions. The imputed ASEC data are coding too many people to be uninsured. The situation is complicated by the current survey items in the ASEC instrument

allowing all members of a household to be assigned coverage with the single press of a button. The Census Bureau should consider altering its imputation specifications and, more importantly, altering how it collects survey data from those who respond to the supplement.

Implications for Policy Delivery or Practice. The bias affects many different policy simulations, policy evaluations and federal funding allocations that rely on the CPS-ASEC data.

Primary Funding Source. The Robert Wood Johnson Foundation.

Key Words. Health insurance coverage, current population survey, annual social and economic supplement, hotdeck imputation, item nonresponse, missing data

---

The U.S. Census Bureau's Annual Social and Economic Supplement (ASEC) to the Current Population Survey (CPS) provides the most visible estimate of the number of uninsured people in the United States. The ASEC has become the survey of record for estimates of health insurance coverage because it produces both national and state estimates of health insurance coverage, makes its micro data available to analysts within 6 months after the data are collected, contains a wealth of demographic information (including family structure and income), and releases its detailed report on an annual basis (Blewett et al. 2004). The ASEC estimates of health insurance coverage are widely used in academic research literature and media outlets, and it is the survey to which all other surveys are compared for coverage measurement.

The ASEC estimates of health insurance coverage are used for a variety of purposes. The Congressional Budget Office makes use of the ASEC to help score legislation (Glied, Remler, and Zivin 2002), and states use the ASEC to monitor progress in determining the success of the State Children's Health Insurance Program (SCHIP) in reducing the number of low income uninsured children in each state (Davern, Blewett et al. 2003). The ASEC is also used to allocate three to four billion dollars per year to states to fund SCHIP based, in part, on the number of low income uninsured children in each state and the number of low-income children in each state (Davern, Blewett et al. 2003).

Address correspondence to Michael Davern, Ph.D., Assistant Professor, Division of Health Policy and Management, University of Minnesota, School of Public Health, 2221 University Avenue, S.E., Minneapolis, MN. Holly Rodin, Ph.D., is with the Service Employees International Union, St. Paul, MN. Lynn A. Blewett, Ph.D., Associate Professor, and Kathleen Thiede Call, Ph.D., Associate Professor, are with the Division of Health Policy and Management, University of Minnesota, School of Public Health, Minneapolis, MN.

Because of their many uses, the ASEC uninsurance estimates have been scrutinized over the years, especially as they tend to be higher than most surveys that ask about health insurance coverage at a single point in time (Lewis, Ellwood, and Czajka 1998; Fronstin 2000; Short 2001; Congressional Budget Office 2003). Some of the national surveys that measure health insurance coverage include the National Health Interview Survey (NHIS), the Household Component of the Medical Expenditure Panel Survey (MEPS-HC), and the Survey of Income and Program Participation (SIPP) (Blewett et al. 2004). The ASEC estimate of the number of people with no health insurance for the entire previous calendar year, is typically higher than the full-year uninsurance estimates produced using data from these other surveys. In addition, the ASEC full-year uninsurance estimates are even higher than many "point-in-time" uninsurance estimates from other surveys (Congressional Budget Office 2003; Czajka 2005; Peterson 2005). Several authors have offered potential reasons why the health insurance estimates differ across the various federal surveys, including: differences in sample frame; sample selection and population coverage; mode of survey administration; survey operationalization of the concept of uninsurance; misreporting by respondents in the survey; and data processing (Lewis, Ellwood, and Czajka 1998; Fronstin 2000; Short 2001; Congressional Budget Office 2003).

One of the least explored reasons why the ASEC differs from other surveys is the impact of missing data imputation on health insurance coverage estimates from the ASEC. Davern et al. (2004) explicitly examined this issue with respect to state survey estimates of health insurance coverage, finding that the process used by the Census Bureau to impute missing insurance data in the ASEC biases the state estimates of uninsurance. Some states had higher and some states had lower rates of coverage due to bias, but the national rate was unbiased (Davern et al. 2004).

In this paper we focus explicitly on the national estimate of uninsurance to explore whether the imputation process used by the Census Bureau explains the higher uninsurance estimates found in the ASEC relative to other national surveys. Specifically, we examine whether there is a significant difference in estimates of uninsurance for those cases that have health insurance data imputed (11 percent of the ASEC sample) and those that do not. We begin by describing the imputation methodology the Census Bureau uses for the ASEC health insurance items and why we think the current methodology may impact the overall rates of coverage to produce an upward bias in the national estimates of uninsurance.

ASEC MISSING DATA AND HEALTH INSURANCE COVERAGE MEASURES

The CPS is a monthly survey conducted by the Census Bureau and sponsored by the U.S. Bureau of Labor Statistics. The ASEC supplement is added to the general CPS monthly survey in February through April of each year. The CPS has a rotating panel design in which sampled households are in the sample for 4 months, then rotate out for 8 months, and then rotate back in for an additional 4 months. The major exception to this rotation is that November respondents in their eighth month of interviewing may be re-interviewed if they meet certain criteria (children or minority members in the household). They are recontacted to take the CPS and ASEC supplement again between February and March (U.S. Census Bureau 2002; Davern, Beebe et al. 2003).

Missing data in the form of item nonresponse is a common problem in survey research (Groves et al. 2002). Missing data results when someone refuses to answer a survey item, an entire supplement, or an entire survey. Item missing data also results when respondents "do not know" an answer to a question. An estimated 11 percent of the CPS monthly core survey respondents do not respond to the entire ASEC supplement. These 11 percent have their entire ASEC supplement values imputed. We refer to these cases as "full-supplement imputations" and they are the focus of our analysis.[1]

Statisticians have developed a wide range of techniques for dealing with item nonresponse (e.g., Kalton 1983; Kalton and Kasprzyk 1986; Little and Rubin 1987; Rubin 1996; Heeringa, Little, and Raghunathan 2002; Marker, Judkins, and Winglee 2001). Most of the techniques use information from the completed cases to impute a model-based estimate to the cases with missing data. The Census Bureau uses "hotdeck" imputation to replace item nonresponse in all of its household surveys (e.g., CPS, SIPP, decennial census, and the Survey of Program Dynamics). Hotdeck is a type of model-based imputation by which a respondent's valid value for a specific variable is assigned to another respondent that does not have a valid value for the variable. The respondent with the valid value is called a "donor" and the one with a missing value is called a "recipient." For example, if the donor is 15 years old, then the recipient (respondent with missing age) is given a value of 15 and the donor maintains the age of 15. Donors and recipients are matched together based on key demographic characteristics that they share in common; for example, sex, age, and work status (for additional information see David et al. 1986; Mason,

Lesser, and Traugott 2001; Marker, Judkins, and Winglee 2001; Davern et al. 2004).

Although properly specified imputation can alter basic distributional summary statistics (means and variances) compared with the statistics calculated using complete cases only, it should not transform the relationships among variables. If there was a relationship between two variables in the reported data it should remain the same in the imputed data, and no new relationships should appear after the imputation. The basic idea of model-based imputation is to use the existing relationships within the reported data to make an informed guess as to what the actual response would have been. In assessing accuracy of datasets with imputed hotdeck values for estimating demographic characteristics, a critical question is whether there is a relationship between having an imputed value and the concept being imputed after controlling for the covariates used to impute the data. If this relationship is strong after controlling for the variables used in the imputation process, there is evidence that bias may have been introduced. To examine whether this is occurring in the ASEC data we explore three questions: (1) is there a difference between the imputed cases and the nonimputed cases with respect to key demographic variables and health insurance coverage?; (2) is there a difference in the probability of being uninsured after controlling for covariates used to impute health insurance coverage?; and (3) is the difference in the probability of being uninsured between those cases with and without imputed data enough to produce a substantively significant difference in the national uninsurance estimate?

WHY IMPUTED CASES MAY BE DIFFERENT FROM REPORTED CASES

The ASEC imputation specification for private insurance coverage has two stages. The first is to impute whether each person is a private health insurance policyholder (done separately for self-purchased and employer sponsored insurance). The second part of the private health insurance imputation process is to impute whether the policyholder has a "family plan" or an "individual plan." If the person is imputed to have a family plan, then dependent coverage is extended to other specific family members. If the person is imputed to have an individual plan, then no dependent coverage is extended to other family members in the household.

The private insurance coverage imputation for the ASEC data has two limitations that we believe are working together to create a significant problem. First, the imputation specification does not use family size when imputing whether or not a person is a private coverage policyholder (U.S. Census Bureau 1998). This is a problem because one-person families are more likely to be private health insurance policyholders after controlling for other covariates. As a result, we expect too few policyholders to be imputed in one-person families, and therefore too little private health insurance coverage among people living in one-person families.

After the private insurance coverage policyholder is imputed, then family or individual coverage is imputed. When a person is imputed to be a policyholder with family coverage, the specification then extends this private coverage to specific relatives living in the same household. Dependent coverage in the imputation process is not extended to a nonchild or nonspouse of the policyholder (U.S. Census Bureau 1998). Although theoretically correct, this is problematic because this limitation is not enforced in the computer-assisted ASEC survey instrument where dependent coverage can be extended to everyone within the household regardless of relationship. Specifically, respondents are asked if "anyone else in the household is covered by this type of insurance" with one acceptable response being "everyone," at which point the Census Bureau's field representative can enter "A" (for "all") regardless of the relationship among the people being assigned coverage. Therefore some nondependent household members may inappropriately be assigned dependent coverage during the survey interview. Because of this difference between the survey interview and the imputation specifications, we expect less dependent health insurance coverage in the imputed data than in the reported data.

ANALYTIC APPROACH

We use the 2004 ASEC supplement to the CPS for this analysis. Data were collected from 77,149 households representing 213,241 individuals; the household response rate for the monthly portion of the CPS was 84 percent (U.S. Census Bureau 2004). In this paper we limit our analysis to ASEC respondents under the age of 65; those above 65 are very likely to have coverage through Medicare.

We examine the full-supplement imputation cases to determine whether there is bias in the national estimates of uninsurance. The first analysis shows

general demographic characteristics of all people in the ASEC under the age of 65. We conduct two independent sample *t*-tests comparing the demographic characteristics of the full-supplement imputations to the nonfull-supplement imputations. The key variables of interest are health insurance coverage as well as other demographic characteristics. The second analysis shows the percent of households in which everyone in the household has coverage by household size. This analysis is further broken into full-supplement imputation households—defined as a household in which any one person in the household is a full-supplement imputation—versus all others.[2]

The third analysis uses a multinomial logistic regression model with the dependent variable taking on one of three values indicating whether the person was coded: (1) to be uninsured; (2) to have any public coverage; and (3) as having private coverage only. If the respondent indicates both private and public insurance, the person is coded as having public insurance.[3] The key covariate of interest in this model is whether the person was a full-supplement imputation.

We use the following variables in our model as they are used in the Census Bureau's health insurance coverage hotdeck routine: age, veteran status, employment, earnings, employer size, self-employment, family labor force participation, and poverty status. In our model we also used several other covariates including race, ethnicity, education, family size, and citizenship in an attempt to see if they are able to explain away the relationship between being an imputed case the probability of being uninsured.

In addition to these key covariates of coverage, we include an interaction term for being under 19 years of age and being a full-supplement imputation. It is possible the impact will vary by whether the person is a child or an adult because children are much more likely to obtain dependent coverage than to be a policyholder. Several of the variables are not collected for children, and we have coded children to not be married, be out of the labor force, and have less than a high school degree.

In our final analysis we estimate the impact of the full-supplement imputation on the uninsurance rate, private coverage rate, and public coverage rate. The impact is assessed in two distinct ways. First, in order to give the multinomial logistic regression results a meaningful scale we use the recycled predictions methodology to obtain estimated coverage rates (Graubard and Korn 1999; StataCorp 2001; Kronick and Gilmer 2002). The recycled predictions approach uses the actual values for a respondent (e.g., black, male, three-person family) to determine the probability of having the various coverage types (e.g., public or private) under a counterfactual scenario. We

alter only one variable in the counterfactual scenario to observe its marginal impact on an individual's probability of coverage. We sum these new altered counterfactual person probabilities for various types of coverage to get the adjusted overall rate of coverage under the scenario. The counterfactual re-cycled predictions analysis gives everyone the value of "not full-supplement imputation." These recycled rates allow us to control for everything else we include in our model (e.g., demographic characteristics and key covariates) while trying to isolate the effect of being a full-supplement imputation. We use this analysis to answer the central question: holding other covariates constant, what is the impact of full-supplement imputation on coverage rates?

We then assess the impact of the full-supplement imputations on the rate of uninsurance and the rate of private health insurance coverage by reweighting the ASEC data after removing all of the full-supplement imputation cases. We adjust the nonfull-supplement imputation cases in the ASEC to population control totals from the entire ASEC by race, ethnicity, gender, age, and poverty status. This analysis helps to further answer the question of what would happen to the ASEC data if the full-supplement cases were treated as non-respondents, as opposed to having their ASEC items fully imputed.[4] We compare the reweighted and recycled predictions estimates to the standard ASEC estimates to gauge the impact of imputation on the health insurance coverage measures.

RESULTS

Table 1 shows the basic demographics of those cases with full-supplement imputations compared with everyone else in the ASEC sample. The full-supplement imputations make up 10.8 percent of the ASEC data and they are significantly less likely to have private insurance coverage: 59.3 percent compared with 69.1 percent for all other cases. There is no significant difference in public coverage between the full-supplement imputations compared with all other cases.[5] The full-supplement imputations also have significantly higher rates of uninsurance, at 26.4 percent compared with 16.6 percent for all others. Among other contrasts, the full-supplement imputations are also less likely to be under 19 years of age, less likely to be working, more likely to be black and less likely to be white, and are more likely to be out of the labor force.

Table 2 compares health insurance coverage rates for those in households where any one person in the household is a full-supplement imputation and those in households without any full-supplement imputations. As expect-

Table 1: Insurance Status and Other Demographic Characteristics of Persons <65 Years of Age by Full Supplement Imputation Status

Variables	Nonfull Supplement Imputations	Full Supplement Imputations	All Cases
Insurance status			
Private coverage	69.1%	59.3%***	68.1%
Public coverage	14.3%	14.4%	14.3%
Uninsured	16.6%	26.4%***	17.6%
Sex			
Female	50.2%	49.8%	50.2%
Male	49.8%	50.3%	49.9%
Age			
≤ 18 years	31.1%	26.8%***	30.6%
19–24 years	9.1%	11.4%***	9.4%
25–34 years	15.4%	16.2%	15.5%
35–44 years	17.1%	17.6%	17.2%
45–54 years	16.1%	17.0%*	16.2%
55–64 years	11.2%	11.2%	11.2%
Marital status			
Married	40.3%	39.2%*	40.2%
Widowed	1.2%	1.3%	1.2%
Divorced	7.5%	7.7%	7.5%
Separated	1.6%	1.8%	1.7%
Never married	49.4%	50.0%	49.4%
Primary family member			
Not family member	39.5%	47.2%***	40.3%
Primary family member	60.5%	52.8%***	59.7%
Family size			
1 person	14.2%	16.6%***	14.5%
2 people	20.0%	21.9%***	20.2%
3 people	19.4%	19.0%	19.4%
4 people	24.3%	21.6%***	24.0%
5+ people	22.0%	20.8%	21.9%
Education			
<High school degree	19.2%	19.9%	19.3%
High school graduate	28.6%	31.2%***	28.9%
Some college	19.0%	19.7%	19.1%
College or postgraduate	33.1%	29.2%***	32.7%
Health status			
Fair/poor	8.1%	8.9%**	8.2%
Excellent/very good/good	91.9%	91.1%**	91.8%
Citizenship			
Not U.S. born	12.0%	12.8%	12.0%
U.S. born	88.0%	87.2%	88.0%
Race/ethnicity			
White only	80.3%	73.8%***	79.6%
Black only	12.5%	18.1%***	13.1%
American Indian	0.9%	0.7%	0.8%

Table 1. *Continued*

Variables	Nonfull Supplement Imputations	Full Supplement Imputations	All Cases
Asian	4.4%	5.4%**	4.5%
Mixed race	1.9%	2.1%	1.9%
Hispanic	15.4%	13.2%***	15.1%
Employment status			
Working	69.5%	65.1%***	69.0%
Unemployed	4.6%	4.2%	4.6%
Out of labor force	25.9%	30.7%***	26.4%
Poverty level			
Under 50%	5.8%	6.7%***	5.9%
50–99%	7.1%	6.6%	7.1%
100–199%	17.3%	16.7%	17.2%
200–299%	16.6%	17.5%	16.7%
300+%	53.3%	52.5%	53.2%
Total percentage and counts			
Unweighted observations	173,386	19,486	192,872
Weighted observations	226,255,311	27,365,896	253,621,207
Percent	89.2%	10.8%	100%

Source: 2004 Current Population Survey—Annual Social and Economic Supplement
*$p<.05$
**$p<.01$
***$p<.001$

ed, the findings are consistent across household size; households with at least one person with a full-supplement imputation have lower rates of coverage. For example, a person living alone who is a full-supplement imputation is

Table 2: Percent of Persons <65 Years of Age in Households Where Everyone Has Health Insurance Coverage by Whether Anyone in the Household Is a Full Supplement Imputation

Household Size	Households with No Full Supplement Imputations (%)	Households with at Least One Full Supplement Imputation (%)	Difference (%)	% Difference between No Full Supplement Imputation Households (%)
1 person	84.6	74.6	10.0	11.8***
2 people	76.2	57.8	18.3	24.0***
3 people	73.4	58.7	14.7	20.1***
4 people	75.5	63.5	12.0	15.9***
5+ people	63.1	53.1	10.0	15.8***

Source: 2004 Current Population Survey—Annual Social and Economic Supplement
***$p<.001$

10 percent more likely to be uninsured; those living in two-person full-supplement households are 18.3 percent more likely to have at least one uninsured households member; and so forth. The percent difference between the full-supplement imputations and the nonfull-supplement imputations is lowest for one-person households, highest for two- and three-person households, with four- and five-person households falling in the middle.

Table 3 shows the results from the multinomial logistic regression model. The coefficient for the full-supplement imputation cases is a strong predictor of whether the respondent is uninsured versus having private health insurance coverage. The full-supplement imputation cases are 2.2 times more likely to have private health insurance coverage relative to being uninsured after controlling for other important covariates. In addition, the interaction effect representing those full-supplement imputations (relative risk ratio = 0.67) from one-person families shows that the impact of being a full-supplement imputation is somewhat reduced for this group. This is because they are not impacted by the survey process that allows all household members to have the same coverage because there are no other household members.

The public insurance coverage estimates were also significant. Full-supplement imputation cases were more likely to be uninsured than to have public insurance coverage. The relative risk ratio was approximately 1.69 and significant, and for full-supplement children it was 0.65 and significant. We believe that full-supplement imputation cases are much more likely to be uninsured because fewer cases are being imputed to have private health insurance coverage. Because the multinomial logistic regression contains interaction effects whose coefficients are difficult to interpret in isolation, we performed additional analyses.

In Table 4 we present the unadjusted ASEC estimates of private coverage, public coverage, and uninsurance for people under 65 years of age. We also break this table into children under age 19 and adults aged 19–64 and report the results from two adjustment methods. The first adjustment is the reweighted data and the second are the "model-based" recycled prediction estimates. Both of these techniques (reweighting and recycled predictions) yield strikingly similar results. The national unadjusted ASEC estimate of people with private coverage is 68.1 percent for those under 65 years of age. In the adjusted "model-based" recycled predictions the estimate is higher, at 69.0 percent, and the reweighted results are slightly higher, at 69.1 percent. The rates for any public insurance coverage do not vary among the three methods. The unadjusted ASEC national estimate of uninsurance is 17.6 percent,

Table 3: Multinomial Logistic Regression Relative Risk Ratios (RRR) for Having Private Coverage versus Being Uninsured and Having Public Coverage versus Being Uninsured for Persons <65 Years of Age

Independent Variables	Private versus Uninsured		Public versus Uninsured	
	RRR	Standard Error	RRR	Standard Error
Full supplement imputation	2.20	0.10***	1.69	0.11***
Full supplement imputation × one person family	0.67	0.06***	0.93	0.11
Full supplement imputation × under 19 years	1.05	0.07	0.65	0.05***
Male	1.17	0.02***	0.99	0.02
Excellent to good health	0.73	0.03***	0.23	0.01***
Black only	1.42	0.05***	1.72	0.07***
American Indian only	1.92	0.21***	1.90	0.20***
Asian only	1.12	0.07	1.07	0.10
Mixed race	1.25	0.10**	1.58	0.11***
Not hispanic	0.48	0.02***	0.59	0.02***
Born U.S. citizen	0.50	0.02***	1.11	0.06*
Employed	1.17	0.04***	0.34	0.01***
Unemployed/out of labor force	2.36	0.12***	0.91	0.06
<50% FPL	10.12	0.56***	24.85	1.46***
50–99% FPL	8.98	0.48***	27.38	1.48***
100–199% FPL	4.43	0.15***	8.62	0.34***
200–299% FPL	1.98	0.07***	2.89	0.13***
Married	0.41	0.01***	0.32	0.01***
One-person family	0.89	0.04*	0.47	0.03***
Two-person family	1.47	0.06***	1.17	0.06**
Three-person family	1.09	0.05*	1.08	0.05
Four-person family	0.86	0.04***	0.82	0.04***
0–18	0.25	0.02***	0.39	0.03***
19–24	1.84	0.09***	0.85	0.05**
25–34	2.00	0.09***	1.08	0.06
35–44	1.34	0.06***	0.88	0.05*
45–54	1.20	0.05***	0.85	0.05**
High school graduate or less	0.73	0.03***	0.70	0.03***
Some college	0.48	0.02***	0.49	0.03***
College or post graduate	0.39	0.02***	0.41	0.02***
Veteran	0.84	0.04***	0.84	0.04***
Employer size 25–499	0.48	0.02***	0.48	0.02***
Employer size 500–999	0.37	0.02***	0.37	0.02***
Employer size 1,000+	0.34	0.01***	0.34	0.01***
Self-employed	1.49	0.06***	1.49	0.06***

Source: 2004 Current Population Survey—Annual Social and Economic Supplement

FPL, federal poverty level.

*$p<.05$

**$p<.01$

***$p<.001$

Table 4:   Comparisons of the Reweighted and Model-Based Predictions for Estimates of Health Insurance Coverage for Persons <65 Years of Age— 2003

Estimate Type	Private Insurance Rate (%)	Public Insurance Rate (%)	Uninsured Rate (%)
Total population			
Unadjusted CPS	68.1	14.3	17.6
Reweighted CPS	69.1	14.3	16.6
Model based CPS (if none imputed)	69.0	14.3	16.7
Children under 19 years of age			
Unadjusted CPS	62.1	26.1	11.8
Reweighted CPS	62.8	26.3	10.8
Model based CPS (if none imputed)	62.8	26.3	10.9
Adults 19–64 years of age			
Unadjusted CPS	70.7	9.1	20.2
Reweighted CPS	71.9	9.0	19.1
Model-based CPS (if none imputed)	71.7	9.0	19.2

Source: 2004 Annual Social and Economic Supplement to the Current Population Survey
CPS, current population survey.

compared with 16.7 percent using the "model-based" recycled predictions and 16.6 percent using the reweighted data. The results within each age stratum follow the same pattern observed for the total population—compared with the unadjusted estimates, the re-weighted and recycled prediction estimates are higher for private coverage, lower for uninsurance, and similar for public insurance.

DISCUSSION

We find evidence that the Census Bureau estimates of health insurance coverage are biased with respect to the full-supplement imputations. We estimate that the bias translates to roughly 2.5 million less uninsured or 6 percent of the total number of uninsured for those people under 65 years of age.[6] This magnitude of difference is similar to the adjustment made to the 1999 uninsurance estimate resulting from the addition of a new health insurance verification question (Nelson and Mills 2001).

The goal of imputation is to use available data to make an informed estimate of what the missing value should be. Imputation should not alter or introduce new relationships among the variables (Davern et al. 2004). Our results show that cases with no imputed values differ significantly from the

full-supplement imputation cases with respect to health insurance coverage. The differences persist even after controlling for the variables that were used to impute health insurance coverage. Our results demonstrate that bias in health insurance coverage estimates is introduced by the current imputation specifications used by the Census Bureau.

Following from these results we recommend that the Census Bureau alter its imputation specifications to eliminate this bias. An important first step in this process is to use family size when imputing private health insurance policyholder status. Adjusting the specifications for who is allowed to receive dependent health insurance coverage is much more complicated. Imputations should reflect the reported data and should not be used to enforce rules that are not enforced in the reported data. As currently designed, the ASEC instrument may allow for too much dependent coverage to be reported among nonimputed households with the "A" for "all" possible response to who else in the household is covered under the policy. With one press of a button dependent coverage can be assigned to people in the household who may not be eligible to receive health insurance coverage through a specific plan. It is quite possible that this assigns dependent coverage to people outside of the family health insurance unit who may not, in fact, be eligible for that coverage. To further complicate the picture, there is evidence from administrative data that a fair number of people actually receiving dependent health insurance coverage benefits do not, in fact, qualify (i.e., are not part of the family health insurance unit). Companies such as Ford and Northwest Airlines have investigated their health insurance coverage roles to remove ineligible people, such as older children and ineligible unmarried partners, who had been enrolled in the plan. In both cases about 10 percent of the people obtaining insurance from the companies were found to not be eligible (Appleby 2004; Cummins and Fedor 2005). As a result we think it is important to actually ask who is covered without necessarily restricting dependent coverage to specific family members as the imputation does.

We also think that simply allowing dependent coverage to be assigned to everyone in the household with a press of a button is not a good practice.

The ASEC survey instrument may allow for incorrect reporting of dependent coverage by allowing the "A" option to assign coverage to all household members, but the imputation specification is not the place to fix this problem. In the short run the Census Bureau should fix the imputed data to reflect the reported data by allowing an imputation of "A" just as the reported data does. However, in the long run the Census Bureau should alter its survey instrument to eliminate the "A" for "all" option. This is especially critical

because dependent coverage is of particular concern as the cost of employer-based coverage continues to increase and the offer and take up of dependent coverage continues to decline (Holohan 2003; Gould 2004). To get a more refined estimate of dependent coverage and policyholders, additional prompts are needed to clearly specify dependent coverage in the ASEC questionnaire. Improved data collection will result in more reliable estimates as well as better baseline data from which to impute missing values.

It is widely acknowledged in the research community that data from the ASEC produce estimates of uninsurance for an entire year that are too high (e.g., Congressional Budget Office 2003; Peterson 2005). In this paper we have documented a previously unexplored reason why these estimates may be too high: the Census Bureau's imputation procedure for full-supplement cases. However, the adjusted estimates presented in Table 4 are still much higher than the full-year uninsured estimates observed in other Census Bureau surveys, such as the SIPP (Congressional Budget Office 2003; Peterson 2005). The imputation bias discussed here is not large enough to reconcile the ASEC estimates with other surveys measuring full-year uninsurance rates. However, the adjusted numbers do bring the estimates closer to other estimates by increasing the amount of private coverage and lowering the number of uninsured.

ACKNOWLEDGMENTS

NOTES

1. Another 2–3 percent of the ASEC respondents have one or more of the items in the health insurance series imputed, but we do not explicitly examine these at this

point. An additional 15 percent of respondents have only dependent health insurance (either privately purchased or employer sponsored) marked as imputed. All of these imputed cases have private health insurance coverage (no one is imputed to be uninsured). After investigation, the Housing and Household Economic Statistics Division of the Census Bureau found that these cases were mistakenly marked as imputed. The Census Bureau informed us that the specification will be altered to remove this problem.

2. For the vast majority (85 percent) of people in households with at least one full-supplement imputation person, everyone in the household is a full-supplement imputation.

3. This is done because the full-supplement imputations are much more likely to have both public and private health insurance coverage than those people who respond to the supplement. Putting those with both public and private health insurance coverage into the public insurance category highlights the main issue regarding imputation and private coverage. To fix this problem, the Census Bureau could impute public coverage first and then use the imputed public coverage in the hotdeck to impute private coverage. This will reduce the number of full-supplement imputations with both types of coverage.

4. This is what is done in other surveys, such as the NHIS, to adjust for supplement nonresponse. In the NHIS the sample child and sample adult supplements have sample loss from the household and person portion of the interview as sampled respondents refuse to take them. These supplements weight the responding cases to represent the entire adult and the entire child population. Supplement refusers do not have their full set of supplement data imputed as in the CPS-ASEC (National Center for Health Statistics 2003).

5. See endnote 3.

6. There are estimated to be 253,621,207 million people <65 in the United States in 2004 multiplied by 1 percent is 2.5 million.

REFERENCES

Appleby, J. 2004. "Employers Check Family Ties to Cut Health Care Costs." *USA Today*, April 6.

Blewett, L. A., M. B. Good, K. T. Call, and M. Davern. 2004. "Monitoring the Uninsured: A State Policy Perspective." *Journal of Health Politics, Policy and Law* 29 (1): 107–45.

Congressional Budget Office. 2003. *How Many People Lack Health Insurance and for How Long? Economic and Budget Issue Brief.* Washington, DC: Congressional Budget Office.

Cummins, H. J., and L. Fedor. 2005. "NWA Seeks to Clean Up Health Roles." *Minneapolis Star Tribune*, March 25.

Czajka, J. 2005. "9 Million Fewer Uninsured? Discussion." Presented at the American Enterprise Institute at the Event Entitled "9 Million Less Uninsured?" [accessed on April 8, 2005]. Available at http://www.aei.org/events/filter.all,eventID.1042/summary.asp

Davern, M., L. A. Blewett, B. Bershadsky, and N. Arnold. 2004. "Missing the Mark? Possible Imputation Bias in the Current Populations Survey's State Income and Health Insurance Coverage Estimates." *Journal of Official Statistics* 20 (3): 519–49.

Davern, M., L. A. Blewett, B. Bershadsky, K. T. Call, and T. Rockwood. 2003. "State Variation in SCHIP Allocations: How Much is There, What are Its Sources, and Can It be Reduced." *Inquiry* 40 (2): 184–97.

Davern, M., T. J. Beebe, L. A. Blewett, and K. T. Call. 2003. "Recent Changes to the Current Population Survey: Sample Expansion, Health Insurance Verification and State Health Insurance Coverage Estimates." *Public Opinion Quarterly* 67 (4): 603–26.

David, M., R. J. A. Little, M. E. Samuhel, and R. K. Triest. 1986. "Alternative Methods for CPS Income Imputation." *Journal of the American Statistical Association* 81: 29–41.

Fronstin, P. 2000. *Counting the Uninsured: A Comparison of National Surveys. Employee Benefit Research Institute Issue Brief No. 225.* Washington, DC: The Employee Benefit Research Institute.

Glied, S., D. K. Remler, and J. G. Zivin. 2002. "Inside the Sausage Factory: Improving Estimates of the Effects of Health Insurance Expansion Proposals." *Milbank Quarterly* 80 (4): 603–36.

Gould, E. 2004. *The Chronic Problem of Declining Health Coverage: Employer-provided Health Insurance Falls for Third Consecutive Year.* EPI Issue Brief no. 202. Washington DC: Economic Policy Institute.

Graubard, B. I., and E. L. Korn. 1999. "Predictive Margins with Survey Data." *Biometrics* 55: 652–59.

Groves, R. M., D. A. Dillman, J. L. Eltinge, and R. J. Little, (eds). 2002. *Survey Nonresponse.* New York: Wiley-InterScience.

Heeringa, S. G., R. J. A. Little, and T. E. Raghunathan. 2002. "Mulivariate Imputation of Coarsened Survey Data on Household Wealth." In *Survey Nonresponse*, edited by R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. Little, pp. 357–720. New York: Wiley-InterScience.

Holohan, J. 2003. *Changes in Employer-Sponsored Health Insurance Coverage. Snapshots of America's Families III, No. 9.* Washington, DC: The Urban Institute.

Kalton, G. 1983. "Compensating for Missing Survey Data." Research Report Series. Ann Arbor, MI: University of Michigan Institute for Social Research.

Kalton, G., and D. Kasprzyk. 1986. "The Treatment of Missing Survey Data." *Survey Methodology* 12: 1–16.

Kronick, R., and T. Gilmer. 2002. "Insuring Low-Income Adults: Does Public Coverage Crowd Out Private?" *Health Affairs* 21 (1): 225–39.

Lewis, K., M. Elwood, and J. Czajka. 1998. *Counting the Uninsured: A Review of the Literature.* Washington, DC: The Urban Institute.

Little, R. J. A., and D. B. Rubin. 1987. *Statistical Analysis with Missing Data.* New York: John Wiley.

Marker, D. A., D. R. Judkins, and M. Winglee. 2001. "Large-Scale Imputation for Complex Survey." In *Survey Nonresponse*, edited by R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. Little, pp. 329–42. New York: Wiley-InterScience.

Mason, R., V. Lesser, and M. W. Traugott. 2001. "The Effect of Item Nonresponse on Nonresponse Error and Inference." In *Survey Nonresponse*, edited by R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. Little, pp. 149–62. New York: Wiley-InterScience.

National Center for Health Statistics. 2003. *National Health Interview Survey (NHIS) Public Use Data Release—NHIS Survey Description.* Hyattsville, MD: Centers for Disease Control and Prevention.

Nelson, C. T., and R. J. Mills. 2001. *The March CPS Health Insurance Verification Question and Its Effect on Estimates of the Uninsured.* Washington, DC: U.S. Census Bureau.

Peterson, C. 2005. "Survey Estimates of the Uninsured and of Medicaid/SCHIP Enrollees." Presentation at the American Enterprise Institute at the Event, "9 Million Less Uninsured?" Washington, DC, [accessed on April 8, 2005]. Available at http://www.aei.org/events/filter.all,eventID.1042/summary.asp

Rubin, D. B. 1996. "Multiple Imputation after 18+ Years." *Journal of the American Statistical Association* 91 (434): 473–89.

Short, P. F. 2001. "Counting and Characterizing the Uninsured." Working Paper Series. Ann Arbor, MI: Economic Research Initiative on the Uninsured, December 2001.

StataCorp. 2001. *Stata Statistical Software: Release 7.0.* College Station, TX: STATA Corporation.

U.S. Census Bureau. 1998. *Health Insurance Edit and Imputation Specifications for the Current Population Survey.* Washington, DC: U.S. Census Bureau.

———. 2002. *Technical Paper 63RV, Current Population Survey: Design and Methodology.* Washington, DC: U.S. Department of Commerce.

———. 2004. *Current Population Survey, 2004 Annual Social and Economic (ASEC) Supplement Documentation.* Washington, DC: U.S. Bureau of the Census.