WORKING PAPER

Today or Last Year?

How Do Interviewees Answer the CPS Health Insurance Questions

JEANNE S. RINGEL
JACOB ALEX KLERMAN

RAND LABOR AND POPULATION

Today or Last Year? How Do Interviewees Answer the CPS Health Insurance Questions?

Jeanne S. Ringel, Jacob Alex Klerman

Jeanne S. Ringel
Economist
RAND Corporation
1776 Main Street
Santa Monica , CA 90407

Phone: 310-393-0411 x6626
Fax: 310-260-8160
E-mail: ringel@rand.org

Jacob Alex Klerman
Senior Economist
RAND Corporation
1776 Main Street
Santa Monica , CA 90407

Phone:310-393-0411 x6289
Fax:310-
E-mail: Jacob_Klerman@rand.org

Abstract

CPS estimates of annual health insurance coverage are below estimates from other surveys and administrative data. One potential explanation is that respondents misinterpret the question and report current rather than past year status. Here we use CPS data matched to administrative enrollment records for Medi-Cal in California to evaluate this possibility. The individual-level matched data allow us to explore the validity of CPS responses under different enrollment scenarios, where "truth" is based on administrative data. While we find some support for the reference period argument, our findings appear more consistent with a combination of cognitive perspectives on interviewing and stigma.

The March Current Population Survey (CPS) provides the most widely cited estimates of the annual health insurance coverage rates. However, the coverage rates generated from the CPS often differ from those generated by other surveys. In general, CPS estimates of annual health insurance coverage are below estimates from surveys such as the Survey of Income and Program Participation (SIPP). In addition, it is well known that CPS estimates of Medicaid coverage are substantially below administrative enrollment totals.[1-5]

A number of potential explanations for the under-reporting of health insurance coverage in the CPS have been suggested. Among the most frequently cited,[7-11] is what we will refer to as "the Swartz conjecture." This explanation was initially put forth in an article by Swartz[6] where she argued that CPS respondents ignore the question wording that specifically asks about any coverage in the previous calendar year, and instead answer the question as if it referred to coverage as of the survey date. This mis-interpretation of the question (i.e., coverage as of the survey) will yield a lower estimate of coverage rates than the question as worded (i.e., coverage last year) as long

as there is some change in health insurance coverage within a calendar year.

In general, it is difficult to test the Swartz conjecture because it is difficult to ascertain the truth against which to compare the CPS responses. In this paper we use CPS data matched to administrative enrollment records for Medi-Cal in California to consider the Swartz conjecture directly. The individual-level matched data allow us to explore the validity of CPS responses under different enrollment scenarios (e.g., enrolled at some point last year, but not in the survey month), where "truth" is based on program administrative data. While we find some support for the Swartz conjecture, the estimates suggest that it is unlikely to be a substantively important explanation for the divergence in estimates of coverage levels across surveys. Overall, our findings appear more consistent with an explanation involving a combination of cognitive perspectives on interviewing and stigma.

The Swartz Conjecture

Swartz[6] conjectured that the lower CPS estimates for health insurance at any time in the calendar year might be due to how interviewees answer the health insurance

questions. The CPS specifically asks about coverage at any time during the previous calendar year:

> At any time in [19XX/20XX] (last year), (were you/was anyone in this household) covered by Medicaid/(fill state name), the government assistance program that pays for health care?
>
> Who was that?

Swartz argued that instead interviewees respond in terms of their coverage status as of the interview date.

There are at least two ways in which this hypothesis might explain why CPS coverage rates are lower than those from other surveys and administrative totals. First, assume that administrative records reflect the truth and that other surveys do a better job of eliciting true annual coverage levels. Other surveys may be better because they more strongly emphasize the intended annual survey response period (by asking both about coverage today and also about coverage in the previous year; e.g., the California Health Interview Survey-CHIS) or because they build up annual coverage estimates explicitly from monthly responses (e.g., the Survey of Income and Program Participation—SIPP).

Second, simple logic implies that as long as there is some change in health insurance coverage over a calendar

year, coverage at a point in time (as Swartz conjectures that respondents interpret the CPS questions) must be lower than coverage over a calendar year (as is implied by the CPS question wording). The divergence between the two measures will be larger when there are more coverage changes within the year. In the case of the CPS, this ordering is not logically neccessary—the point in time response is not included in the calendar year reference period. However, assuming that coverage levels are relatively stable and there is enough change in coverage from month to month, the ordering remains empirically plausible and coverage rates in March will be lower than coverage at any time in the previous year.

Matching CPS Responses to Administrative Data

To allow providers to verify eligibility of individuals for covered services, the California Department of Health Services operates the Medi-Cal Eligibility Data System (MEDS). The MEDS is a single, statewide, individual-level record of Medi-Cal eligibility in each month. The data are verified by the state and used for the provision of services, so they are believed to be of relatively high quality. Therefore, for this analysis, we treat the MEDS information as "truth."

The MEDS uses Social Security Number (SSN) as its primary individual identifier. The CPS attempts to collect SSNs from every respondent age 14 or older (hereafter, we refer to them as adults). Some respondents refuse to provide SSNs and others give what the Social Security Administration deems an invalid SSN.

Through the Census Bureau's Research Data Center program, we matched California CPS respondents with a valid SSN to their corresponding MEDS record, where one existed (those who had never received Medi-Cal would not match to any MEDS record). To protect the confidentiality of CPS respondents, the match was done in the UCLA site of the California Census Research Data Center on files pre-processed by Census such that all SSNs were replaced with Cenus constructed pseudo-SSNs. See Klerman, Ringel, and Roth[12] for details of the file construction and sample selection criteria.

Our analysis file includes CPS surveys for 1991, 1994, and 1997-2000. For this analysis we choose to focus on only the highest quality data. Specifically, we analyze records for individuals who provided an SSN, for whom the Social Security Administration validated the SSN, and if there was a match for whom gender and age (plus or minus one year)

were consistent across the CPS and MEDS data. During this period, each CPS interview had about 50,000 households. About a tenth of them were in California and we lose about a third of the records for lack of a valid SSN or a bad match. The final sample includes about 56,300 people.

This Month vs. Last Year

These data allow us to examine how the CPS response varies with actual patterns of Medi-Cal enrollment. Exhibit 1 explores the Swartz conjecture directly. It divides our matched sample into four groups by their MEDS enrollment status in the past year and in March of the following year and tabulates the CPS response. As expected, there is significant under-reporting of Medi-Cal coverage in this sample. According to the CPS instructions, people who are covered in the past year (the first two rows of Exhibit 1) should answer in the affirmative. In this sample, this group represents 10.8 percent (=8.7%+2.1%) of the total. However, based on CPS responses in the full sample, only 9.5 percent actually respond in the affirmative. Thus in aggregate, CPS estimates of Medi-Cal coverage reflect only 88 percent of the true level. (See Klerman, Ringel, and Roth,[12] who show that under-reporting is even more common in the unmatched sample.)

Focusing on specific patterns of enrollment and CPS response sheds light on the question of interest. The people who have the same status in both last year and in the interview months (shown in the first and last rows of Exhibit 1) are not informative for the Swartz conjecture. However, we note that most, but far from all (79 percent), of those with Medi-Cal last year and in the following March (first row) respond in the affirmative. Conversely, few (2 percent) of those with Medi-Cal neither last year nor the following March (last row) respond in the affirmative. Put differently, the error rate is higher among those with Medi-Cal, than among those without. Even so, because the group that was not enrolled last year or in the survey month is so large (89% of the sample), the actual number of reporting errors attributed to this group is relatively high.

To evaluate the Swartz conjecture, we focus our attention on the people for whom the status last year diverges from the status at the interview (the second and third rows of Exhibit 1). They are the only cases that are informative about whether misinterpretation of the reference period affects survey response. The CPS question explicitly asks about coverage at any point during the previous year, so in the second row of Exhibit 1 the

7

percentage of people responding "Yes" in the CPS "should be" 100 percent and the third row "should be" zero. Alternatively, the Swartz conjecture implies the opposite responses. As can be seen in Exhibit 1, neither simple explanation explains all of the responses.

Among those enrolled last year, but not this month (i.e., the second row), the pattern of responses is consistent with the Swartz conjecture. Within this group, 27 percent give the correct "Yes" response reflecting past year enrollment status. At the same time, the other 73 percent report status consistent with their enrollment during the survey month. Among those enrolled this month, but not last year (i.e., the third row), however, the pattern of reporting does not support the Swartz conjecture. Within this group, 84 percent give the correct "No" response while only 16 percent give a "Yes" response consistent with their current enrollment status.

Note that, as would be expected for a stationary stochastic process, the group of people who were enrolled last year but not this month (2.1% of the sample) is much larger than the group enrolled this month but not at any point last year (0.5% of the sample). Moreover, the responses within this larger group are generally consistent

8

with the Swartz conjecture. Thus, again consistent with the Swartz conjecture, among all people for whom last year and this month are not consistent, this month is the more likely response.

However, rather than interpreting these responses as being consistent either with the question as worded (i.e., last year) or with Swartz conjecture (i.e., this month), these results seem to be more consistent with a stigma-based explanation. It appears that people want to report not being enrolled in Medi-Cal and tend to do so when they have any cognitive reason to justify the response (e.g., interpreting the question in such a way that they can respond negatively). For both groups, the "No" response is more common (73% among those enrolled last year, but not this month and 84% among those not enrolled last year, but enrolled this month).

Finally, it is important to note that at least for Medi-Cal in our sample, the Swartz conjecture is unlikely to explain much of the net under-reporting. Enrollment status for the past year and survey month diverges for only a small number of people (2.7 percent of the sample). Assuming that all people who were on Medi-Cal last year and not in the survey month who misreport coverage do so

because they misinterpret the reference period explains only 48 percent of the false negatives. Thus, at the very most, misinterpretation of the time frame can explain less than half of the under-reporting.

Instead, most of the under-reporting is among those who were enrolled last year and this month (first row of Exhibit 1). The CPS estimate of coverage for this group is only 79% of the true level translates to an under-reporting rate of 21%. This rate is much higher than the 12% under-reporting rate for the full sample. This divergence is to a great extent caused by the offsetting false positives. As noted above, among those with Medi-Cal neither last year, nor this month, reporting errors are rare, about 2 percent. However, that group is so large (89 percent of the full sample) that this small rate makes up a large share of all of the Medi-Cal responses in the CPS, 21 percent (=1.99/9.50)

The Dose-Response Relationship

Looking at the number of months of Medi-Cal coverage in the past year provides additional insight into the questions of interest. From the MEDS data, we compute months of enrollment last year and compare this measure across people who were enrolled in the survey month and

10

those who were not. The observation counts at each month are too small to allow reporting the raw rates. Instead Exhibit 2 reports raw rates for (1) those enrolled in March of this year, and all of last year; and (2) for those not enrolled in March of this year and none of last year. For the other combinations, we report the results of logit polynomial regressions. Reporting the regression results also smoothes away some of the sampling variation.

If reporting were perfect, the points for "0" months enrolled would be zero and the other points would be 100 percent. Neither the number of months (if greater than zero) covered last year, nor enrollment status in March of this year "should " matter. Instead, we observe a clear dose-response relationship. Regardless of March enrollment status, the more months of enrollment last year recorded in the MEDS, the more likely a person is to report program enrollment in the CPS. In addition, enrollment in March affects the probability of answering the CPS question about last year positively. These findings are consistent with standard cognitive approaches to survey response[13-14] that suggest that an affirmative Medi-Cal response is likely to be more salient and therefore positively related to both the number of months enrolled in the past year and enrollment status in the survey month.

11

Focusing on those people enrolled at least one month in the prior year, we see that overall the effect of March enrollment on the probability of falsely reporting no coverage (one minus the percent reporting Medi-Cal enrollment in the CPS) grows as the number of months enrolled increases. In other words, enrollment in the month of the survey can explain a larger share of the false negatives among those people enrolled for most of last year than for those people only enrolled for a small number of months.

Discussion

This paper exploits CPS responses on Medi-Cal coverage matched to California administrative data on "true" Medi-Cal enrollment to improve our understanding of the response errors in the CPS. While our results provide some support for Swartz's conjecture that people answer the CPS as though it was a point in time survey, we believe that a more appropriate interpretation involves a combination of cognitive perspectives on interviewing and stigma.

More broadly, these tabulations emphasize the magnitude of the response errors in the CPS. Even people enrolled all of last year and in the survey month only report Medi-Cal about 85 percent of the time. Those

12

enrolled in the survey month and a single month last year only report Medi-Cal about 30 percent of the time. While only a small percentage of those with Medi-Cal neither last year nor last month answer that they have Medi-Cal, this group is so large as to induce a larger number of false positives.

References

1. Lewis, K., Ellwood, M., and Czajka, J.L. (1998). *Counting the uninsured: A review of the literature.* Washington, DC: The Urban Institute.

2. Coder, J, and Ruggles, P. (1988) "Welfare Recipiency as Observed in the SIPP." *SIPPWorking Paper Series* No. 8818, U.S. Bureau of the Census, Washington, D.C.

3. Ku, L. and Bruen, B. (1999) "The Continuing Decline in Medicaid Coverage," *New Federalism: Issues and Options for States,* Series A, No. A-37, Washington, DC: The Urban Institute.

4. Bavier, R. (1999). An Early Look at the Effects of Welfare Reform. Unpublished manuscript, Office of Management and Budget, Washington, DC.

5. Bavier, R. (2003). Non-Economic Factors in Early Welfare Caseload Declines. Unpublished manuscript, Office of Management and Budget, Washington, DC.

6. Swartz, K. (1986). Interpreting the Estimates from Four National Surveys of the Number of People without Health Insurance, *Journal of Economic and Social Measurement* 14(3):233-42.

7. Bilheimer, L. T. (1997). CBO Testimony on Proposals to Expand Health Coverage for Children. Testimony before the Subcommittee on Health. U.S. House of Representatives, Committee on Ways and Means, Washington, DC, April 8.

8. Buchmueller, T. and Valletta, R. (1999). The Effect of Health Insurance on Married Female Labor Supply. *Journal of Human Resources,* 34(1), Summer, pp. 42–70.

9. Carrasquillo, O., Himmelstein, D.U.,Woolhandler, S. and Bor, D. H.. (1999). A Reappraisal of Private Employers' Role in Providing Health Insurance, *The New England Journal of Medicine*, 340(2):109-114.

10. Carrasquillo, O., Carrasquillo, A. I., and Shea, S. (2002). Health Insurance Coverage of Immigrants Living in the United States: Differences by Citizenship Status and Country of Origin. *American Journal of Public Health,* 90(6), June.

11. Kronebusch, K. (2002). Children's Medicaid Enrollment: The Impacts of Mandates, Welfare Reform,

and Policy Delinking. *Journal of Health Politics, Policy and Law*, 26(6).

12.   Klerman, J. A., Ringel, J. S., and Roth, E. A. *Under-Reporting of Medicaid and Welfare in the Current Population Survey.* RAND/WR-169-3, 2005.

13.   Sudman, S. and Bradburn, N. (1973) "Effects of Time and Memory Factors on Response in Surveys." *Journal of the American Statistical Association,* 68: 805-815.

14.   Groves, R. M. (1989). Survey Errors and Survey Costs. New York, NY: Wiley and Sons.

Exhibit 1 - CPS Reference Period

MEDS			CPS		
Last	Survey				%
Year	Month	Row %	CPS='Yes'	CPS='No'	Yes
Y	Y	8.65%	6.85%	1.80%	79%
Y	N	2.14%	0.58%	1.56%	27%
N	Y	0.51%	0.08%	0.43%	16%
N	N	88.70%	1.99%	86.71%	2%
Total		100.0%	9.50%	90.50%	10%

Notes: First two columns give MEDS status ("Last Year"—
any enrollment in the previous calendar year; "Survey
Month"—enrollment in the March of the current year).
Columns labeled "CPS='Yes'" and "CPS='No'" give the
unconditional probability of enrollment in Medi-
Cal/Welfare. Column labeled "% Yes" gives the
probability of a "Yes" response in the CPS given the
MEDS data for the survey month and the previous
calendar year (i.e., within the row).

Exhibit 2 — CPS Reporting of Medi-Cal Given MEDS Pattern of Receipt