The Influence of State-Specific Programmatic Characteristics in the Modeling of Medicaid Undercount;

A Record-Check Study of the 2001 Current Population Survey Annual Social and Economic Supplement (CPS ASEC)

Matthew Stevens
US Census Bureau
4600 Silver Hill Rd. Washington, DC 20233 / matthew.j.stevens@census.gov

Any views expressed on statistical, methodological, and technical issues are those of the author and not necessarily those of the U.S. Census Bureau

Abstract
Much research has been conducted regarding underreporting of benefit receipt on population surveys. Specifically, much work has been devoted to measuring Medicaid undercount on the Current Population Survey Annual Social and Economic Supplement (CPS ASEC). Previous studies have revealed estimates of false-negative reporting rates varying from 10-36%. As one of the primary indicators of Medicaid benefit receipt, measurement error within this estimate will not only bias survey results, but subsequently bias the program evaluation, policy development, and funding decisions informed by the CPS ASEC. Upon gaining estimates of the size of underreporting, the next round of research focused on seeking to explain why Medicaid recipients provide erroneous reports. Record check studies conducted by the Census Bureau revealed significant associations between respondent-level characteristics (such as recency and intensity of coverage, as well as household income and relationship to person being reported for) and underreporting.

Using the same administrative records database utilized in previous studies, this study searches for the root of state-level differences in underreporting that still exist after controlling for respondent-level characteristics. Specifically, how effectively can the characteristics of state-run Medicaid programs (such as eligibility rules, policy on premiums, and portrayal to the public) explain the likelihood of recipients within in said state to fail-to-report benefits? Examination of Medicaid programmatic literature enabled creation of a series of indicator variables to represent these state programmatic differences within logistic regression models. Results suggest that while many characteristics of the respondent effect Medicaid reporting, characteristics of the state Medicaid program itself may equally impact reporting propensity, and that such sources of measurement error may be easier to remedy on the CPS ASEC than previously identified, respondent specific sources.

Introduction
Policymakers and foundations rely on survey estimates to provide a glimpse of the uninsured and benefit recipients. While these estimates fall short in their attempts to capture the entire population of recipients, they are a necessity when discussing programs lacking a centralized database of participants. Prior to 1999, Medicaid, a state-run program providing health coverage and long-term care support services, lacked a centralized database. The Balanced Budget Act (BBA) of 1997 changed this by requiring all states to submit participant-level data quarterly to the Centers for Medicare and Medicaid Services (CMS), thus creating a centralized administrative records database: the Medicaid Statistical Information System (MSIS).

With the creation of such a database, the quality of survey estimates regarding public participation rates could finally be examined at a national level. Consistent with the notion of imperfect survey estimates, prior state-level studies had revealed a consistent undercount of Medicaid reporting (estimates varying from 10-30%) (Card et al, 2001; Czajka & Lewis, 1999; Blumberg and Cynamon, 1999; Lewis et al, 1998; Klerman, Ringel and Roth, 2005). These studies not only revealed the presence of measurement error (in the form of undercount) in prior surveys of Medicaid participation, but patterns of non-randomness within this error. Respondents with certain demographic and Medicaid-related characteristics were more or less likely to provide false-negative reporting (claim to not be receiving Medicaid coverage when the administrative records database proved that they were covered) than respondents with other characteristics. With aspirations of both establishing the true Medicaid undercount as well as examining the relationship between false-negative reporting and respondent attributes, the University of Minnesota's State Health Access Data Assistance Center, the Centers for Medicare and Medicaid Services, the Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation, and the US Census Bureau (this collaborative deemed SNACC based on the first letters of each participating organization) worked in collaboration on the four-phase "Research Project to Understand the Medicaid Undercount."

The task central to this project involved the linkage of Current Population Survey's Annual Social and Economic Supplement (CPS ASEC) responses (from years 2000 and 2001) to Medicaid administrative records data in the MSIS. After steps are taken to remove the known difference between the data sets and comparability is ensured, this research was able to attribute almost all of the 36.2% Medicaid undercount to measurement error (SNACC Phases I & II). Consistent with the results of previous smaller, state-specific analyses, this measurement error was found to be highly non-random. Specifically, levels of recency and intensity of Medicaid coverage (Pascalle et al 2008) as well as the income level and relationship to the person being reported for (Lynch 2008) were all determined to be significant predictors of false-negative Medicaid reporting. The CPS ASEC serves as one of the primary indicators of Medicaid benefit receipt, and thus it is important that the estimates are not only accurate, but that they are free from measurement error to ensure accurate representation of participants within the results. Measurement error with regards to program participation not only biases survey results, but also subsequently biases program evaluation, policy development, and funding decisions informed by the CPS ASEC. Therefore, it is essential to not only reduce measurement error, but to explore and understand the patterns of error that exist.

Construct validity is a difficultly detected but prominently present source of measurement error in surveys. The underlying motivation for survey administration is to extract information (through questions) in an attempt to map the collective of individual, personal experiences to the sum of these experiences, thereby capturing the pulse of society as a whole. The ability to attain this underlying goal is limited to the extent to which the information being provided by the respondent matches that which the question intends. Traditionally, discussion of construct validity has been limited to questions regarding abstract concepts (Peter 1981; Andrews 1984), but cross-cultural survey theory suggests that socio-geographic context can alter the construct validity of even the most straight-forward, concrete concepts. For instance, Verba (1971) suggests that a concept as simple as that of fruit can elicit very different imagery depending what fruits are traditionally available in the vicinity of a respondent. With this in mind, we would expect the respondent's perception of the construct of Medicaid to vary based on the characteristics of the program in his/her state.

The aforementioned studies of CPS Medicaid underreporting have revealed significant patterns in measurement error. Characteristics of the respondent and his/her Medicaid history create a unique reality from which the reports of coverage are based. Variation in these realities should be expected to produce variation in propensity to recognize, much less report, benefit receipt. As a state-run, federally funded program, the experience of Medicaid receipt will be unique for respondents in different states. Additionally, we would expect these environmental differences to impact the manner in which respondent specific characteristics lead to measurement error. This paper will decompose the individual-level characteristics from the state-level characteristics related to a respondent's Medicaid experience to answer the following question: can these state-specific factors further explain the underreporting amongst the population of CPS respondents receiving benefits? First, literature on Medicaid policy will be examined to identify programmatic characteristics known to vary between states. Next, upon controlling for the recipient-level factors identified in the Lynch (2008) paper, we will use a series of logistic regression models to capture the remaining effect of state-level programmatic factors. Finally, we will interpret the role of state-level factors within these models to make recommendations for reducing underreporting on the CPS ASEC.

Methods

Highlighting Medicaid underreporting at this theoretical level, we create a format that will be used to describe measurement error for the remainder of the paper. Figure One below describes the propensity for underreporting as the synthesis of both individual-specific and state-specific factors. 'Y' is set to either '0' (report receipt) or '1' (fail to report receipt, i.e. under-report/false-negative response). Modeling Y through logistic regression, under-reporting propensity can be decomposed into two main aspects: (1) those factors specific to the individual regarding his/her demographic characteristics as well as his/her experience with Medicaid, and (2) factors specific to the state Medicaid program which the respondent is enrolled in.

Figure One: Model to Decompose the Roots of False-Negative Reporting

$$\hat{Y} = \beta_0 + \beta_{1,i} x_i + \beta_{2,j} x_j$$

where…

$\hat{Y}$: propensity to underreport Medicaid coverage

β_0 : intercept of logistic regression equation

$\beta_{1,i}$: effect of factor "i", a factor regarding the respondent and his/her Medicaid experience

x_i : the level for variable "i" for a given respondent

$\beta_{2,j}$: the effect of factor "j", a factor regarding the state Medicaid program in which the respondent is enrolled

x_j : the level for variable "j" for a respondent

Prior research (Lynch 2008) took a similar approach to modeling the underreporting of Medicaid receipt. Using individual-level variables such as age, minority status, income, as well as a series of variables to represent recent medical and medical benefit history, her study was able to identify the factors that significantly represent the $ß_1$ aspect of our model. Her models also included indicator variables for 50 states and the District of Columbia that would represent the $ß_2$ portion of the model above. Finding state-differences but no discernible patterns with strong enough significance to use 'state' as a predictor (within her analysis) justifies the need to look deeper at the inherent state-level programmatic differences behind these effects.

Figure One outlines the conceptual basis for the models used in this paper. Using logistic regression, we will predict propensity to underreport as a combination of these respondent specific ($ß_{1,i}$'s) and state-specific ($ß_{2,j}$'s) variables. Within this notation $ß_{1,i}* x_i$ represents the summation of the products of all respondent-specific variable values multiplied with the appropriate regression coefficients. The $ß_{1,j}* x_j$, on the other hand, represents the summation of the products of all state-specific variables values multiplied with the appropriate regression coefficients. While more thorough notation would include all 29 $ß_{1,i}$'s and 6 $ß_{2,j}$'s, such a summary would place too much emphasis on the predictors themselves and not enough to the overlying concepts they represent. Though more detail will be given to these specific predictors later in the paper, Figure One exists merely to provide a framework in which to group (and more importantly, think about) our predictors.

As state-funded and implemented entities receiving additional federal funding, the characteristics of specific Medicaid programs change from year to year with budgetary constraints. As the faces of these programs change, so do the experiences of the program participants. Therefore, even though the CPS ASEC asks the same Medicaid question to respondents from all states, the programs that respondents are enrolled in may be quite different from one another, and thus the process of providing an accurate response may be quite different. Specifically, there are three main attributes of state Medicaid programs in which variation occurs: program eligibility rules, policy on charging premiums, and program portrayal to the public (Ross and Marks 2009).

With regards to eligibility rules, some states include presumptive eligibility for children, providing immediate but temporary access to Medicaid benefits. In order to maintain benefits, completion of the full application process is required within a month of initial receipt. Respondents covered under presumptive eligibility have therefore experienced both a medical incident on a previously non-covered child, as well as completed a full-application process within the immediate period. One would expect such experiences to be more salient for respondents than those receiving standard Medicaid coverage.

Enrollment procedures vary with eligibility requirements from state to state. Specifically, many states offer 12-month continuous eligibility: guaranteeing uninterrupted eligibility for families with children for 12 months regardless of changes in income or other circumstances that may otherwise compromise eligibility (Broaddus et al., 2003). In addition to being guaranteed eligibility, families are not required to re-apply for coverage until the 12-month period has elapsed. Recipients in such states therefore experience less fluctuation in coverage and fewer experiences with the administrative aspects of Medicaid, which often serve a construct for an otherwise face-less entity. Our theory would therefore suggest less salience for recipients residing in states with these characteristics.

The last aspect of program eligibility that will be examined concerns the maximum income in which a family can still receive Medicaid benefits. Within our data set, this variable is expressed as maximum income with respect to the federal poverty line. All states provide coverage to low-income families. States with higher cutoff levels will simply be providing coverage to a more heterogeneous pool of recipients with a wider range of income levels. Therefore, any social stigma that may be related to receiving public benefits would expect to be reduced. With fewer stigmas, we would expect to see lower level of underreporting in these states.

Policy on charging premiums for coverage varies from state to state and is the second attribute in our explanation of state-level trends in underreporting. Beginning in the 1990s, states were allowed to apply for section 1115 waivers. Introduced to encourage states to implement premium assistance within their Medicaid programs, these waivers relax benefit, cost-sharing, and cost effectiveness requirements that were previously mandatory for states wishing to receive federal funding (Alker 2005). Medicaid recipients in states with such waivers are less likely to pay premiums for coverage. Our theory of salience of participation suggests that the cost of paying for coverage would increase the salience, especially amongst lower-income respondents.

Public portrayal is the third category of variation in state-level Medicaid underreporting. Literature suggests that confusion surrounding the name of Medicaid programs is one the sources contributing to the Medicaid undercount (Kincheloe et al., 2006). As some states have state-specific names for their programs other than "Medicaid" it will be interesting to examine under-reporting in these states compared to the remainder of states. Intensity of media effort in Medicaid promotion will serve as the other indicator of program portrayal to the public. We would expect to see lower under-reporting rates in states that have more intensive media programs (those utilizing more than one media format), as Medicaid would be more salient to CPS respondents within such states.

Building on the research of SNACC to identify both the presence and the roots of the Medicaid undercount, this paper explores the programmatic differences at the state-level as well as their contribution to the differences in state-level undercount. First, the state-level trends in underreporting will be examined in order to gain a better understanding of the scope of variation. Second, indicator variables will be coded to represent the state-level, programmatic contribution to underreporting propensity. Next, a model developed through prior research from the Census Bureau will be run to re-identify the respondent-specific attributes related to underreporting. Controlling for these individual-level predictors, the state-level programmatic indicators will be used to model the remaining variation in underreporting. Finally, examination of the coefficients on these indicator variables will provide insight on the sources of measurement error in the CPS ASEC Medicaid question, and recommendations will be made on how to create a less error-prone measurement of Medicaid receipt.

Data Resources

This paper builds on the findings of the Phase II SNACC project, and therefore the analysis file is constructed around the pre-existing file. Phase I of the SNACC project focused on the creation of a national database of Medicaid recipients by first evaluating the quality of records contained within the MSIS administrative records file. These records were run through the Person Verification System (PVS), which in a two iteration process first validates the Social Security Number (SSN) and then in a separate process replaces each SSN with a Protected Identification Key (PIK). This PIK will eventually be utilized to match these records to survey data while maintaining the anonymity of the individual respondents (SNACC I Documentation).

For the next step, efforts were made to remove respondents from the MSIS who may have been living in institutionalized group quarters and thus outside of the 2001 CPS-eligible universe. For the remaining MSIS records, "months enrolled in Medicaid" were determined and output to the MSIS Summarized Enrollment History File (MSEHF). The final MSEHF consisted of 39,911,501 person-records: one record for each unique PIK, each record containing information highlighting the entirety of Medicaid receipt.

CPS ASEC records underwent a similar process, preparing them for accurate linkage. First, the 218,269 sampled person records were run through the PVS to attain anonymity and assign PIK. In about 20% of the persons, SSN were either unknown or unverifiable. Such cases were deemed unusable and dropped from the analysis file. The weights of dropped CPS cases were reassigned proportionally to retained records within the same re-weighting strata. The authors of the SNACC Phase II report admit that this re-weighting is likely to introduce bias into the data, since valid SSN's are more-than-likely not missing at random, but nonetheless, they deem it a necessary move in order to make the results of their study represent the entire CPS frame. Finally, CPS (173,967 records) and MSEHF (the 39,911,591 processed MSIS records) files are linked via PIK. This final analysis data set contains one record for every person in CPS. Since we are interested in false-negative reporting versus true-positive reporting, we will be looking exclusively at the 23,643 person records in which we have MSIS information.

Having established how the analysis file is constructed, some mention should be made regarding the CPS ASEC, as well as what Medicaid information it contains. While the Current Population Survey (CPS) is a monthly face-to-face survey that collects data regarding labor force participation amongst non-institutionalized civilians living in the United States, the Annual Social and Economic Supplement (ASEC) is conducted once a year (usually in March) to collect information specific to work, income, public assistance, and health insurance in the previous year. Integral to this analysis are seven questions relating to health insurance. Following three questions on private coverage (employer-sponsored, directly-purchased, and coverage from someone outside the household) are questions on four government-related plans (Medicare, Medicaid, SCHIP, and military plans). For this analysis, a false-negative response is any person for whom a MSIS record exists but who did not have a "yes" response to the CPS ASEC Medicaid question. Therefore, the dataset used for analysis will consist of the 23,643 Medicaid recipients, their Medicaid information from the MSIS, socioeconomic and demographic information from the CPS, as well as a flag to represent whether or not they reported receipt of benefits.

Results

Before understanding the roots of state-level differences in Medicaid underreporting, we must first examine underreporting levels on a state-by-state basis. Figure 2 (in the Appendix) provides the details of the underreporting rate broken out by state for the 23,643 respondents included in the analysis. As described in the methods section, these respondents represent CPS ASEC respondents who were successfully matched to the MSIS Medicaid database; therefore all are true Medicaid recipients. The rate of underreporting is expressed as the ratio of persons in our data set whom responded "no" to the Medicaid question over the total number of persons from that state. It should be noted that these estimates are unweighted, as we hope to draw conclusions on the individual sample respondents rather than make inferences about the entire Medicaid population within the CPS. With this in mind, when discussing state-levels of underreporting rate, we are simply referring to the underreporting rates in a given state amongst CPS respondents who were successfully matched to persons known to receive Medicaid benefits (those on the MSIS). In so much as the intention of this paper is to understand patterns of underreporting within the universe of matchable CPS respondents, this should be viewed less as limitation and more as an assumption for interpretation.

Prior research from the Census Bureau (Lynch 2008) used the same MSIS-matched CPS respondent data set to examine the factors that were related to Medicaid underreporting. This research found significant relationships between respondent characteristics and underreporting. These findings reiterated the nonrandom nature of underreporting amongst CPS and suggested that after controlling for several demographic and Medicaid programmatic variables, significant differences still existed between states. Figure 2 (Appendix) supports these findings, as much variation in underreporting rates exists between states. While CPS states have a mean rate of 43.63% false-negative reporting (std. error of 6.94%) for the people examined, state rates vary from 23.54% (Rhode Island) on the low-end to 63.45% (Hawaii) on the high-end.

Figure 3 is a geospatial display of this distribution: the darkest green areas represent the states with the lowest underreporting rates and the darkest red areas represent the states with the highest underreporting rates. There appears to be a slight correlation between geographic proximity and underreporting amongst CPS respondents, as states in similar ranges are clustered or adjacent to one another. Based on our aforementioned theories, we would expect proximal states to experience similar levels of underreporting to the degree that proximal states have socio-political similarities, which in turn, would create similar Medicaid programs.

Figure Three: Geospatial Display of Medicaid Underrporting Rate

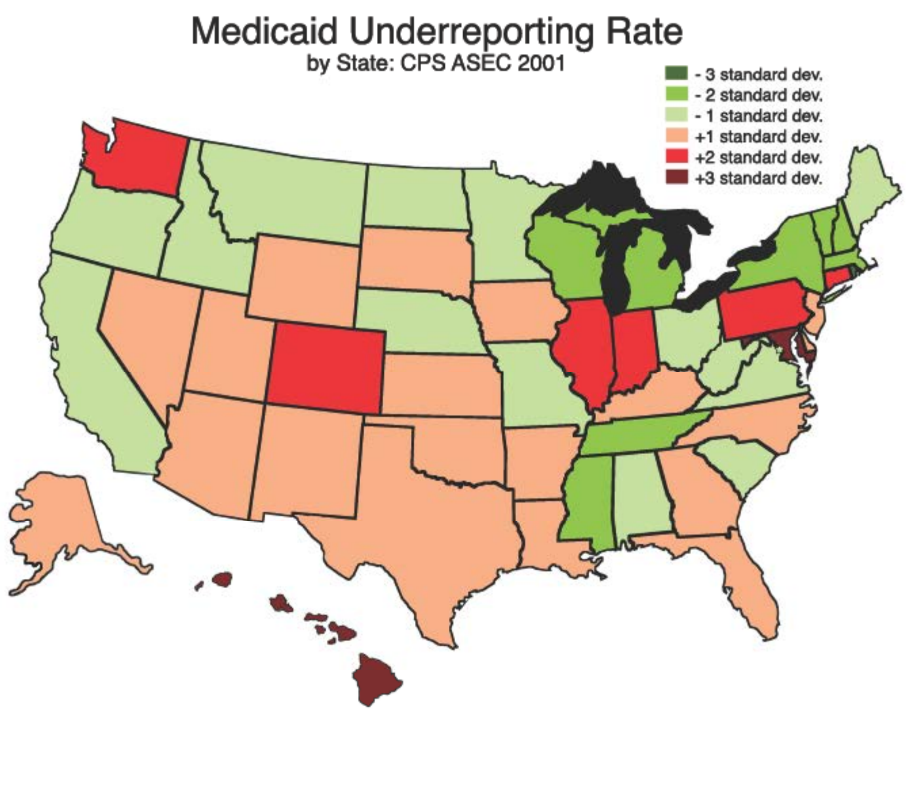

Figure 4 (Appendix) shows how states were coded within the research indicator variables. These maps provide the reader with a bit more insight about the distribution of these indicator variables. Some of the geospatial clusters that can be seen in figure three reappear within these maps, further illuminating the relationships between these programmatic indicators and underreporting. Bivariate relationships provide a first glimpse at the relationships between these concepts and underreporting. For states with 12-month continuous eligibility, higher levels of underreporting can be seen (43.18% compared to 41.90%, which falls just a tad short from being significantly different at the 10% level). Looking at the final column of figure five, we can see that our other four binary indicator variables all present differences that are statistically significant at the 5% level. States with 1115 waivers, alternative program names, and that offer presumptive eligibility for children all appear to have decreased underreporting rates: illustrated by lower values in final column amongst the 'yes' row than the 'no'. 1115 waivers are often characterized by the decreased likelihood of having some premium. This finding is not consistent with the theory that the process of paying would increase the salience of Medicaid program participation. As the Figure 5 (Appendix) indicates, states with 1115 waivers have an underreporting rate of 38.10% compared to a rate of 44.05% in all other states. As described previously, presumptive eligibility for children is a program characteristic that can grant children immediate, temporary access to Medicaid benefits. In order to maintain benefits, the child is required to complete the full application within a month. This negative relationship (38.52% for states with presumptive eligibility compared to 43.35% in other states) is consistent with our theory that such cases would lead to more salient experiences for the participants. Alternative program name, on the other hand, has results that are not consistent with the expectations of this study. While it was anticipated that underreporting would be higher in states whose Medicaid programs have alternative names, the opposite effect can be seen in Figure 5 (Appendix) (42.34% in states with alternative names; 45.38% in all other states). Finally, the variable for multiple media formats appears to be positively correlated with Medicaid underreporting. States who utilized more than one format (television, radio, and/or print) to promote their Medicaid program saw higher rates (43.97%) than those states that devoted their resources to single format (38.04%). Having found significant differences in underreporting rate based on levels of our binary indicator variables, as well as having gained a better understanding of the distribution of both our dependent and independent variables across states, we utilize multivariate logistic regression models to explore the role of state-specific factors within a respondents propensity to provide false-negative reports.

We first replicate the Lynch (2008) Model to capture the respondent-specific, individual-level factors that impact propensity to provide a false-negative response. Figure 6 (Appendix) summarizes the regression coefficients, standard errors, and significance levels for this model in the column marked "Model One/Control Model." This control model verifies many of the findings of the initial study: significant negative relationships for younger reference persons, children of respondents, those respondents receiving both Medicaid and Medicare, those receiving Social Security, and those who have paid for prescription and/or non prescription services in the past 90 days (with stronger effects for those purchases more recently within the 90-day period); significant positive relationships for Hispanic or minority respondents, and other (not spouse, child, self, or parent) relationship to reference person (the respondent). The first column in figure six illustrates how this model utilizes demographic and recipient-specific programmatic variables to provide an explanation of underreporting trends. As the distribution of these variables tends to vary for each state, the fifty-one binary variables representing state had various significant and insignificant values within this model (Note: these values have been omitted from Figure 6 for the sake of space, as their specific values are not as important to our analysis). Therefore these controls should serve to remove the individual-level effects ($ß_1$) from our later models, thereby isolating the contribution of state-level characteristics (ß2) to underreporting propensity. These state-level characteristics will later be modeled by our programmatic indicator variables.

Next, the model is simplified and is re-run using only the indicator variables created to capture the characteristics of the state Medicaid programs. Having already described the relationships between each indicator and underreporting, this logistic regression model reveals how all the indicators work together to explain the dependent variable. Column two of Figure 6 (Appendix) (titled "Model Two/Indicator Model") provides the coefficients for the indicator model. Comparing these values to those in Figure 5 (Appendix), we see that the directions of these indicators are consistent between bivariate and multivariate models. While this serves to further solidify the direction of the effects of our state-characteristics on underreporting, it should be noted that only two of the six variables provide statistically significant coefficients: Multiple Media Formats (0.1184 w/ standard error .0404) and Maximum Income-to-Poverty Ratio for Eligibility (-0.2232 with standard error .0604). Looking towards the bottom of the Model Three column of figure five, we see 12-month continuous eligibility has a coefficient of .0489, but neither this value nor the coefficients for 1115 waivers or Alternative Program names are significantly different from zero. At this point in the modeling process, this lack of significance is not troubling, as we have seen clearly defined patterns within the data. Given the large number of control variables in later equations, it is likely that significance will be hard to attain. Therefore, our conclusions are likely to be constructed from a consolidation of results rather than a single definitive model. With this in mind, we will use the findings of model four to reiterate the conclusions discussed in the following section.

Discussion

The first goal of our research was to replicate the results of the Lynch paper by modeling underreporting propensity with respect to the influential participant-level variables (the $ß_1$ component within our model). Consistent with these results, we see that answering the Medicaid question for any of the following persons significantly reduces underreporting: younger reference persons, reference persons who are the children of the respondent, those respondents receiving multiple public benefits such as Medicare and Social Security, and those who have paid for some medical care in the past 90 days (the more recent the payment, the less likely underreporting of Medicaid benefits). Younger reference persons, particularly a respondent's children, are more likely to be in the direct care of the respondent. Consistent with the findings of Lee et al. 1999, information regarding Medicaid benefits for such reference persons is more likely to have been encoded than for some other related or non-related household member. Additionally, it could be argued that events affecting closer reference persons will be more salient in the mind of the respondent. Significant positive relationship between "reporting for other respondent" (other than spouse, child, self, parent) and underreporting further supports this notion. This increased salience is likely motivating the decreased underreporting for those respondents with recent medical expenses. These findings are all consistent with Census studies performed on this CPS/MSIS data set (Lynch 2008) (Pascale et al).

Before delving into the interpretation of our findings, it is important to reassess our design and address any potential limitations to our results. One limitation that may potentially hinder our ability to detect significant effects is that while a state program has a specific characteristic, there is no guarantee that the chosen CPS respondents have experienced this characteristic. From this perspective, this modeling is more exploratory than explanatory, providing us with a first glimpse as to how our variables impact underreporting when taken simultaneously. Our research was based on the logic that if respondents had the opportunity to be exposed to a characteristic (i.e.. lived in a state with said characteristic) then the impact of that characteristic would be larger (even if just slightly so) in such states. To determine true causality of state-level characteristics, future research would need to exclusively study respondents exposed to state-level predictors (for example, a sample of respondents who benefited from presumptive eligibility for children). Another limitation involves the dataset itself. While record check studies are agreed upon as the most effective way to detect measurement error, the samples available for such studies are limited to those survey respondents who can successfully be matched to records in the administrative records database. The SNACC project utilized Social Security Numbers (SSNs) in its person-matching procedure. Therefore, CPS respondents without a reported SSN will not be matched and subsequently excluded from our analysis. This is an acknowledged limitation of both this research and all research conducted as a part of the SNACC project. Future research is needed to examine the implications of this less-than-ideal data situation.

While comprehension of the participant-level variables is essential to understanding the full scope of underreporting trends, our research focuses on how the respondent behaves within the context of his or her state-specific Medicaid environment ($ß_2$ component within our model). Exclusive analysis of state-specific indicators (Model Two) revealed that the indicators related to eligibility (12-month continuous eligibility, maximum family income, and presumptive eligibility) were all related to the dependent variable in the same direction as expected, while premiums (1115 waivers) and portrayal to the public (alternative program names and multiple media formats) are related in the opposite direction than anticipated.

Figure 7: Summary of Model Two (Indicator Model)

$$\hat{Y} = \beta_0 + \beta_{2,j} x_j$$

or

$$\hat{Y} = \beta_0 + \beta_{2,a,j} x_{aj} + \beta_{2,b,j} x_{aj} + + \beta_{2,c,j} x_{aj}$$

where…

$\hat{Y}$: propensity to underreport Medicaid coverage

β_0 : intercept of logistic regression equation

$\beta_{2,j}$: the effect of factor "j", a factor regarding the state Medicaid program in which the respondent is enrolled

x_j : the level for variable "a, j" for a respondent

$\beta_{2,a,j}$: the effect of factor "a, j", a factor regarding eligibility rules within the state Medicaid program in which the respondent is enrolled

$x_{a,j}$: the level for variable "b, j" for a respondent

$\beta_{2,b,j}$: the effect of factor "b, j", a factor regarding premiums within the state Medicaid program in which the respondent is enrolled

$x_{b,j}$: the level for variable "j" for a respondent

$\beta_{2,c,j}$: the effect of factor "c, j", a factor regarding public portrayal within the state Medicaid program in which the respondent is enrolled

$x_{a,j}$: the level for variable "c, j" for a respondent

While this reduced model provides us with a sense of the general direction of these effects, it possesses no practical interpretable implications on its own. Our theory suggests that while state-specific factors may impact how a respondent interprets and answers the Medicaid question, we fully expect these factors to be secondary to the participant-level factors. That is, those factors at the state-level are merely explaining the differences that exist after controlling for the factors relating to a specific respondent within that state. The full model (Model Four as seen in Figure Five) takes the findings of the previous exploratory model and applies them to our theory. Three out of six of our indicator variables (representing two out of three of our conceptual sources) prove to be significant predictors of Medicaid underreporting. 'Presumptive eligibility for children' and 'maximum family income' are both indicators designed to represent the concept of programmatic/eligibility rules and both are found to have negative relationships with the dependent variable. That is, states with presumptive eligibility and higher maximum incomes for eligibility have lower rates of underreporting and therefore more accurate reporting of Medicaid rates. From a measurement error perspective, these findings are consistent with literature regarding both encoding and reporting of sensitive topics in face-to-face interviews. More interestingly, it allows us to break down the $ß_{2a}$ component into two sub-components: those that elicit purposeful deception and those that spur inadvertent underreporting.

The indicator representing states offering presumptive eligibility for children is consistently found to have a significant negative coefficient within our models. As a reminder, this program provides children with immediate but temporary access to Medicaid benefits. Essentially, if an uninsured child shows up for emergency medical care in one of these states, he/she will be treated and receive support for up to thirty days without having to undergo the full Medicaid application process. While literature (Lee et al. 1999) suggests that parents will be less likely to encode specifics regarding medical procedures received by a child, it is likely that they will encode the experience of receiving unexpectedly affordable health care in an emergency situation. In the case that they then applied for Medicaid for their child, the application process increases saliency, again increasing the probability that they will have encoded this receipt of benefits. To the degree that failure to encode information can account for underreporting of Medicaid receipt, states with presumptive eligibility should expect more accurate results of the CPS question.

The other significant component of the $ß_{2,a}$ component is represented in our model by maximum family income to still be eligible for benefits. This is the aspect of the model influenced by states in which it is socially desirable for some respondents to purposefully fail to report Medicaid receipt. The negative relationship between this variable and underreporting suggests that states in which the maximum family income for eligibility have lower underreporting rates. This relationship is consistent with a multitude of literature regarding the tendency for respondents to provide inaccurate socially desirable answers in such instances. To fully understand this relationship it may be advantageous to examine the other side: states with lower maximum incomes experience higher levels of underreporting. Such states thus experience more economic homogeneity amongst Medicaid recipients and these recipients have across-the-board lower incomes. Therefore, within the social structure of such states, Medicaid is more likely to be viewed as a lower-class benefit, and thus viewed with social stigma. Of course, the CPS instrument cannot hope to affect the level of social stigma, but the designers could take steps to decrease the bias associated with it. In states with lower maximum incomes (or even in all states), the Medicaid question could be self-administered. To the extent that social desirability is responsible, we would expect to see a reduction in underreporting (an increase in reporting accuracy).

The $ß_{2,b}$ portion of our model ends up being wholly insignificant at predicting underreporting. 1115 waivers were designed to represent whether or not participants in the state programs were required to pay premiums for their healthcare. 1115 waivers tend to lower the costs of premiums, and in some cases, eliminate them altogether. Therefore, a negative relationship was expected: the payment of higher premiums would be more salient than the act of not paying at all. Payment would essentially provide an additional opportunity for the respondent to encode their enrollment and use of Medicaid. From the cognitive response perspective, perhaps payment and eligibility are deeply intertwined. In that case, the significant impact on saliency of coverage inspired by presumptive eligibility may completely account for this effect.

The final aspect of our model ($\beta_{2,c}$) represents those aspects of Medicaid programs that are visible to the public and the respondent but do not have anything to do with the substance of the programs. The indicator to represent those states in which Medicaid is called something different was not found to be a significant predictor of underreporting. In fact, this variable was the only one to switch positive/negative signs between different models, further suggesting a regression coefficient not significantly different from zero. These findings do not support our hypothesis that Medicaid programs referred to as such ("Medicaid") would experience less underreporting. The other indicator in this section represents those states that utilized more than one form of media (amongst print, television, and radio) to make the public aware of their services. Originally, it was theorized that states utilizing multiple media formats would better reach the public, increasing public awareness of Medicaid and thus reducing underreporting. The results of our analysis contradict this theory. Perhaps media exposure goes beyond making the public aware of a service. It could be argued that in areas where Medicaid is being promoted from multiple sources of media, that this community-shared awareness eventually leads to a shared public perception of benefits. Such a situation could very easily lead to perception that some stigma is associated with Medicaid, thus increasing the possibility that respondents in these states are purposefully providing inaccurate answers. Similar to reducing the error associated with "maximum income for eligibility" providing the respondent a chance to privately self-report Medicaid receipt could prove to be advantageous.

Conclusion

First, we set out to create a framework with which to think about false-positive Medicaid reporting at the respondent level. By isolating the respondent specific characteristics impacting this propensity, we theorize that the remaining contributors vary along state-lines in so much as the Medicaid programs vary from state-to-state. Utilizing literature on Medicaid policy, indicator variables were developed to represent these programmatic state differences. Logistic regression analysis allowed us to add these programmatic variables to preexisting models that relied exclusively on respondent characteristics to predict propensity of false-negative reporting. Amongst the eligibility-related variables, we conclude that the undercount is caused both by purposeful deception (related to the stigma a respondent associates with receiving benefits in the socioeconomic climate of his/her state) as well as inadvertent underreporting (based on some states having programmatic details that create less salient Medicaid experiences for the respondent). Premiums are found to be a wholly insignificant predictor of underreporting. Three interpretations can be made from this finding: 1) the relationship between the increased saliency of paying premiums and reporting receipt of benefits is controlled for by some other variables in our model, 2) a relationship exists between premiums and reporting, but presence of 1115 waivers within a state are not adequate indicators of premium payment, 3) there is in fact no significant relationship between the payment of premiums and reporting. Finally, contradicting our initial expectations, we find that factors thought to raise a respondent' awareness of Medicaid actually increased false-negative reporting. In conjunction with our other findings, this result suggests that as much as the programmatic details can impact a respondent' perception of his/her benefits, underreporting is largely related to the level of social stigma for benefits within a state.

The next steps for research in this field should further decompose the sources of error in the Medicaid undercount. Where this study divided the error sources into two categories (those related to the respondent and those related to the Medicaid program within a state), future research should decompose the state-level analysis into those factors related to salience vs. those factors related to stigma. Salience-based errors can be resolved by tailoring the CPS ASEC survey instrument to better represent the reality of the Medicaid program within that state, while allowing the question to be self-administered can reduce stigma-based errors. Whatever decisions are made in the future regarding the CPS Medicaid question, thorough testing of the instrument could ultimately reduce these errors. With the creation of a continually updated MSIS administrative records database, the results of any testing could be matched to both determine the undercount rate as well as continue to study to underlying patterns and trends that impact false-negative reporting propensity.

Acknowledgements

Special thanks to Amy Ohara for her efforts in granting me access to the data set and to Victoria Lynch for her help providing Medicaid background information as well as invaluable data resources. Thanks to Toni Warner and Leah Marshall for their efforts in reviewing the final version of this paper. Finally, many thanks to Melissa Rodgers for her graphical expertise, as well serving as a constant source of encouragement.

References

Alker, Joan. 2005. "Premium Assistance Programs: How are they financed and do states save money?" Kaiser Commission on Medicaid and the Uninsured. Washington, DC.

Andrews, Frank M. 1984. "Construct Validity and Error Components of Survey Measures: a Structural Modeling Approach." *Public Opinion Quarterly*. Vol. 48. p. 409-442.

Blumberg, Stephen J., Marcie Cynamon. 1999. "Misreporting Medicaid Enrollment: Results of Three Studies Linking Telephone Surveys to State Administrative Records." Proceedings of the Seventh Conference on Health Survey Research Methods, DHHHS Publication No. (PHS) 01-1013. Hyattsville, MD

Broaddus, Matthew, Blanely, Shannon, Dude, Annie, Guyer, Jocelyn, Ku, Leighton, Peterson, Jaia. 2002. "Expanding Family Coverage: States- Medicaid Eligibility Policies for Working Families in the Year 2000." Center on Budget and Policy Priorities. Washington, DC.

Card, David, Andrew K.G. Hildreth, Lara D. Shore-Sheppard. 2001. "The Measurement of Medicaid Coverage in the SIPP: Evidence from California, 1990-1996." National Bureau of Econometric Research, Working Paper 8514. Cambridge, MA.

Czajka, John L., Kimball Lewis. 1999. "Using National Survey Data to Analyze Children's Health Insurance Coverage: An Assessment of Issues," Report submitted by Mathematical Policy Research to the U.S. Department of Health and Human Services. Washington, DC.

Kincheloe, Jennifer, Brown, Richard E., Frates, Janice, Call, Kathleen Thiede, Wei, Yen, Watkins, Jim. 2006. "Can We Trust Population Surveys to Count Medicaid Enrollees and the Uninsured?" *Health Affairs,* Jul/Aug2006, Vol. 25 Issue 4, p. 1163-1167

Klerman, Jacob Alex, Jeanne S. Ringel, Beth Roth. 2005. "Under-Reporting of Medicaid and Welfare in the Current Population Survey." RAND Labor and Population working paper series WR-169-3. Santa Monica, CA.

Lewis, Kimball, Marilyn Ellwood, John Czajka. 1998. "Counting the Uninsured: A Review of the Literature." The Urban Institute. Washington, DC.

Lynch, Victoria. 2008. "Medicaid Enrollment: The Relationships between Survey Design, Enrollee Characteristics, and False-Negative Reporting" US Census Bureau Report. Suitland, MD.

Pascalle, Joanne, Marc I. Roemer, Dean Michael Resnick. 2008. "Medicaid Underreporting in the CPS: Results from a Record Check Study." US Census Bureau Report. Suitland, MD.

Peter, J. Paul., "Construct Validity: A Review of Basic Issues and Marketing Practices." *Journal of Marketing Research.* Vol. 18., No. 2., May 1981. p. 133-145.

Ross, Donna Cohen, Laura Cox. 2004. "Beneath the Surface: Barriers Threaten to Slow Progress on Expanding Health Coverage of Children and Families: A 50 State Update on Eligibility, Enrollment, Renewal, and Cost-Sharing Practices in Medicaid and SCHIP." Kaiser Commission on Medicaid and the Uninsured. Washington, DC.

Ross, Donna Cohen, Caryn Marks. 2009. "Challenges of Providing Health Coverage for Children and Parents in a Recession: A 50 State Update on Eligibility Rules, Enrollment and Renewal Procedures, and Cost-Sharing Practices in Medicaid and SCHIP in 2009." Kaiser Commission on Medicaid and the Uninsured. Washington, DC.

Smith, Vernon, Kathleen Gifford, Eileen Ellis, Robin Rudowitz, Molly O˜Malley, Caryn Marks. 2007. "As Tough Times Wane, States Act to Improve Medicaid Coverage and Quality: Results from a 50-State Medicaid Budget Survey for State Fiscal Years 2007 and 2008." Kaiser Commission on Medicaid and the Uninsured. Washington, DC.

SNACC I (Research Project to Understand the Medicaid Undercount: The University of Minnesota's State Health Access Center, the Centers for Medicare and Medicaid Services, the Department of Health and Human Services Assistant Secretary for Planning and Evaluation, The National Center for Health Statistics, and the US Census Bureau). 2008. "Phase I Research Results: Overview of National Medicare and Medicaid Files."

SNACC II (Research Project to Understand the Medicaid Undercount: The University of Minnesota's State Health Access Center, the Centers for Medicare and Medicaid Services, the Department of Health and Human Services Assistant Secretary for Planning and and Evaluation, The National Center for Health Statistics, and the US Census Bureau). 2008. "Phase II Research Results: Examining Discrepancies between the National Medicaid Statistical Information System (MSIS) and the Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC)."

Stevens, Matthew J. 2009. 'The Influence of State-Specific Programmatic Characteristics in the Modeling of Medicaid Undercount; A Record-Check Study of the 2001 Current Population Survey Annual Social and Economic Supplement (CPS ASEC)." Master's Thesis, Joint Program in Survey Methodology, University of Maryland.

Verba, S., "Cross-National Survey Research: The Problem of Credibility." In: *Comparitive Methodology on Sociology – Essays on Trends and Applications*. Vallier I. (ed.). Berkley: University of California Press.

Willis, G., Brittingham, A., Lee, L., Tourangeau, R., & Ching, P. (1999). "Response errors in surveys of children's immunizations." *Vital and Health Statistics*, Series 6, Number 8. Hyattsville, MD: National Center for Health Statistics.

Appendix

Figure Two: CPS Medicaid False-Negative Reporting Rates by State

	Number of CPS Respondents Failing to Report Medicaid Receipt	Number of CPS Respondents Found on the MSIS to have Received Medicaid Benefits	Percentage of False-Negative Reports of Medicaid Benefit Receipt
Alabama	158	412	38.35%
Alaska	189	380	49.74%
Arizona	169	358	47.21%
Arkansas	142	323	43.96%
California	806	2111	38.18%
Colorado	136	250	54.40%
Connecticut	243	393	61.83%
Delaware	141	301	46.84%
District of Columbia	137	360	38.06%
Florida	495	958	51.67%
Georgia	190	364	52.20%
Hawaii	158	249	63.45%
Idaho	112	282	39.72%
Illinois	451	841	53.63%
Indiana	191	341	56.01%
Iowa	123	276	44.57%
Kansas	105	236	44.49%
Kentucky	145	311	46.62%
Louisiana	164	323	50.77%
Maine	160	457	35.01%
Maryland	147	237	62.03%
Massachusetts	188	585	32.14%
Michigan	174	533	32.65%
Minnesota	135	323	41.80%
Mississippi	107	348	30.75%
Missouri	157	391	40.15%
Montana	71	185	38.38%
Nebraska	127	352	36.08%
Nevada	132	260	50.77%
New Hampshire	73	273	26.74%
New Jersey	213	437	48.74%
New Mexico	228	505	45.15%
New York	447	1304	34.28%
North Carolina	251	547	45.89%
North Dakota	107	252	42.46%
Ohio	238	661	36.01%
Oklahoma	175	365	47.95%
Oregon	113	322	35.09%
Pennsylvania	420	763	55.05%
Rhode Island	117	497	23.54%
South Carolina	157	364	43.13%
South Dakota	151	332	45.48%
Tennessee	177	525	33.71%
Texas	519	1029	50.44%
Utah	135	283	47.70%
Vermont	216	692	31.21%
Virginia	99	247	40.08%
Washington	267	505	52.87%
West Virginia	161	399	40.35%
Wisconsin	114	349	32.66%
Wyoming	114	252	45.24%
Total	10,145	23,643	42.91%

Figure Four: Geospatial Display of Indicator Variables Across States

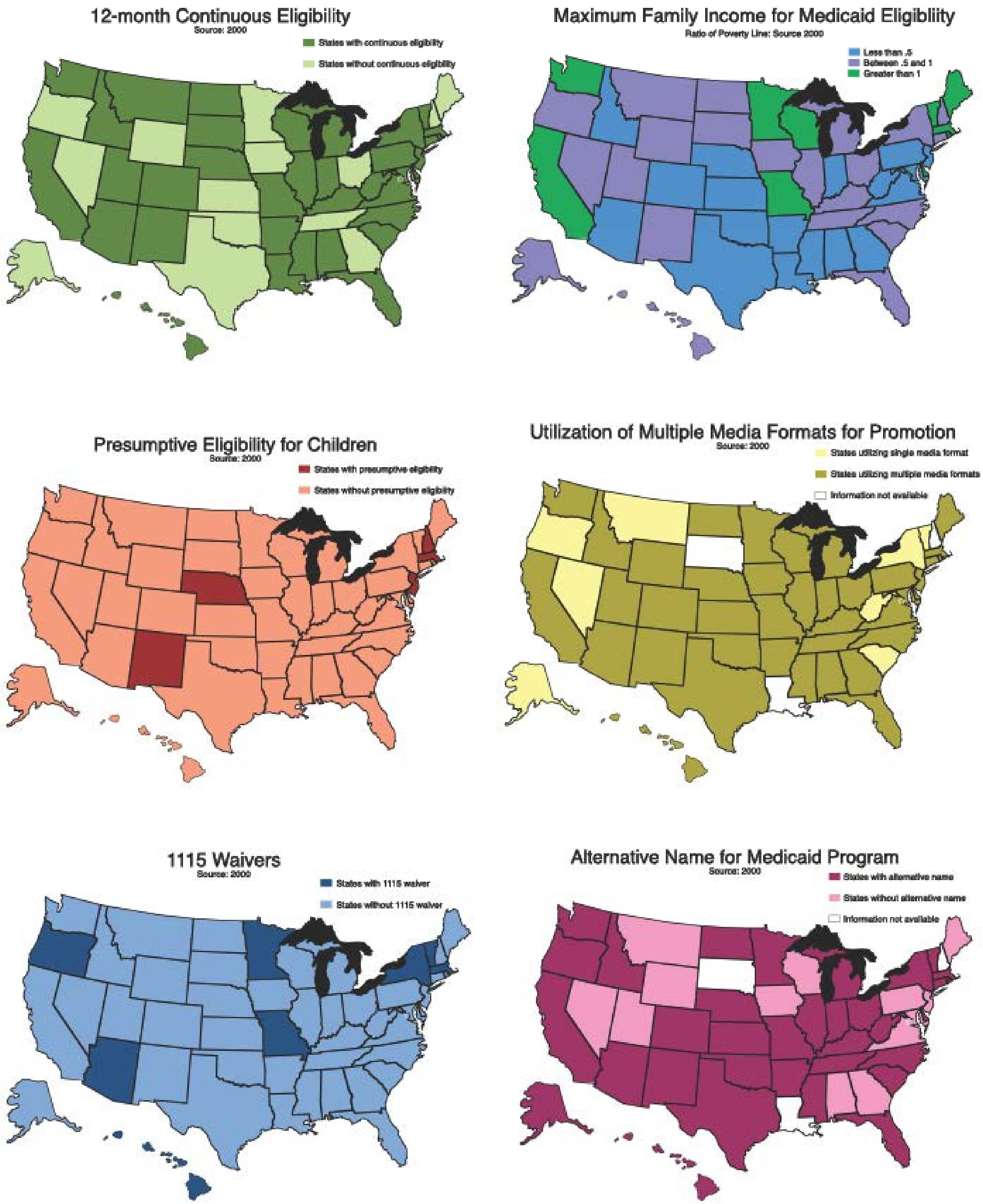

Figure Five: False-Negative Reporting Rate for Respondents Residing in States with Certain Medicaid Characteristics
[* - significant at the .10 level, ** - significant at the .05 level]

		Number of CPS Respondents Failing to Report Medicaid Receipt	Number of CPS Respondents Found on the MSIS to have Received Medicaid Benefits	Percentage of False-Negative Reports of Medicaid Benefit Receipt
12-Month Continuous Eligibility	No	2130	5083	41.90%
	Yes	8015	18560	43.18%
1115 Waiver**	No	8421	19118	44.05%
	Yes	1724	4525	38.10%
Multiple Media Formats**	No	1486	3906	38.04%
	Yes	8271	18809	43.97%
Alternative Program Name**	No	2090	4606	45.38%
	Yes	7667	18109	42.34%
Presumptive Eligibility**	No	9316	21491	43.35%
	Yes	829	2152	38.52%

Figure Six: Summary of Logistic Regression Models to Predict Medicaid False-Negative Reporting [- significant at the .10 level, ** - significant at the .05 level]*

		Model One *Control Model* Coefficient (std. error)	Model Two *Indicator Model* Coefficient (std. error)	Model Three *Indicator Model w/ Controls* Coefficient (std. error)	Model Four *Significant Indicator Model w/ Control* Coefficients (std. error)
Intercept		1.5291** (.0736)	-0.2065** (.0833)	1.4531** (.1210)	1.5089** (.1158)
Age of Reference Person	0-5	-0.3314** (.0618)		-0.3154** (.0617)	-0.3143** (.0617)
	6-14	-0.1734** (.0550)		-0.1751** (.0550)	-0.1745** (.0551)
	15-17	-0.1176* (.0671)		-0.1088 (.0701)	-0.1098 (.0701)
	18-44	0.1088* (.0465)		0.0959** (.0471)	0.0947** (.0471)
	45-64	-0.433 (.0739)		-0.0485 (.0724)	-0.0499 (.0719)
Hispanic or Minority	Yes	0.2279** (.0594)		0.2308** (.0584)	0.2349** (.0585)
Sex	Male	-0.0332 (.0417)		-0.0283 (.0410)	-0.0286 (.0410)
Relationship to Reference Person	Parent	-0.1123 (.1406)		-0.1039 (.1371)	-0.1018 (.1365)
	Spouse	0.0660 (.0767)		0.0627 (.0754)	0.0620 (.0750)
	Own Child	-0.1968** (.0519)		-0.1946** (.0522)	-0.1949** (.0523)
	Other	0.2314** (.0647)		0.2156** (.0642)	0.2149** (.0641)
Income-to-Poverty Ratio	< 50%	-0.5250** (.0622)		-0.5072** (.0608)	-0.5110** (.0607)
	50-74%	-0.4842** (.0769)		-0.4791** (.0778)	-0.4783** (.0778)
	75-99%	-0.1975** (.0588)		-0.1957** (.0574)	-0.1983** (.0573)
	100-124%	-0.1122 (.0718)		-0.0981 (.0733)	-0.0957 (.0731)
	125-149%	0.1157* (.0694)		0.0801 (.0713)	0.0819 (.0718)
	150-174%	0.1308 (.0915)		0.1315 (.0940)	0.1328 (.0943)
	175-199%	0.3727** (.0891)		0.3700** (.0894)	0.3693** (.0898)
Medicaid Supplements Medicaid	Yes	-0.5512** (.1106)		-0.5531** (.1131)	-0.5525** (.1131)
Medicaid with Private Insurance	Yes	-0.00691 (.0966)		-0.00982 (.0979)	-0.00049 (.0980)
Social Security Insurance	Yes	-0.7076** (.0696)		-0.7230** (.0680)	-0.7170** (.0679)
Payments for Medical Service in 2000	Yes	-0.3757** (.0570)		-0.3611** (.0564)	-0.3670** (.0574)
Payment for Prescriptions (past 30 days)	Yes	-0.3482** (.0494)		-0.3624** (.0512)	-0.3624** (.0513)
Payment for Prescriptions (past 31-60 days)	Yes	-0.2896** (.0797)		-0.3025** (.0807)	-0.3030** (.0806)
Payment for Prescriptions (past 61-90 days)	Yes	-0.2337** (.0903)		-0.2607** (.0908)	-0.2648** (.0915)
Payment for Non-Prescription Service (past 30 days)	Yes	-0.7228** (.0593)		-0.7567** (.0595)	-0.7512** (.0596)
Payment for Non-Prescription Service (past 31-60 days)	Yes	-0.6407** (.1030)		-0.6309** (.1062)	-0.6254** (.1056)
Payment for Non-Prescription Service (past 61-90 days)	Yes	-0.3648** (.1261)		-0.3587** (.1267)	-0.3552** (.1267)
Temporary Assistance for Needy Families (TANF)	Yes	-0.3637** (.0741)		-0.3861** (.0762)	-0.3740** (.0761)
51 State-Level Binary Variables		Various Values			
12-Month Continuous Eligibility	Yes		0.0299 (.0357)	0.0489 (.0351)	
1115 Waiver	Yes		-0.0348 (.0378)	-0.0663 (.0427)	
Multiple Media Formats	Yes		0.1184** (.0404)	0.1037** (.0446)	0.1395** (.0399)
Alternative Program Name	Yes		-0.0447 (.0323)	0.00463 (.0357)	
Presumptive Eligibility	Yes		-0.0736 (.0534)	-0.1337** (.0592)	-0.1440** (.0570)
Max Income-to-Poverty Ratio for Elig.			-0.2232** (.0604)	-0.2135** (.0633)	-0.2391** (.0582)