Published in final edited form as:
Stat J IAOS. 2019 August 26; 35(3): 443–456. doi:10.3233/sji-180470.

Multiple imputation to account for linkage ineligibility in the NHANES-CMS Medicaid linked data: General use versus subject specific imputation models

Jennifer Rammon[*], **Yulei He**, **Jennifer D. Parker**
Division of Research and Methodology, National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, MD, USA

Abstract

Data from the National Health and Nutrition Examination Survey (NHANES) have been linked to the Center for Medicare and Medicaid Services' Medicaid Enrollment and Claims Files. As not all survey participants provide sufficient information to be eligible for record linkage, linked data often includes fewer records than the original survey data. This project presents an application of multiple imputation (MI) for handling missing Medicaid/CHIP status due to linkage refusals in linked NHANES-Medicaid data using the linked 1999–2004 NHANES data. By examining multiple outcomes and subgroups among children, the analyses compare the results from a multi-purpose dataset produced from a single MI model to that of individualized MI models. Outcomes examined here include obesity, untreated dental caries, attention deficit hyperactivity disorder (ADHD), and exposure to second hand smoke.

Keywords

Children; data linkage; NHANES; Medicaid; multiple imputation

1. Introduction

The Medicaid program is the largest health insurance program in the United States. Together with the Children's Health Insurance Program (CHIP), Medicaid covers over thirty percent of all children, over fifty percent of low-income children, and over forty percent of all births in the United States [1]. In 2014, children represented 43% of overall Medicaid enrollment and 17% of all Medicaid expenditures [2]. Given that such a large number of children rely on Medicaid and CHIP coverage for their health care, understanding the health status of these enrollees is important. Future assessments of the Medicaid and CHIP program rely on a clear evaluation of the health status of Medicaid and CHIP children.

The National Health and Nutrition Examination Survey (NHANES) provides national estimates from in-home interviews and physical examinations. The NHANES biomarkers

[*]Corresponding author: Jennifer Rammon, Centers for Disease Control and Prevention, National Center for Health Statistics, Division of Research and Methodology, HYAT Bldg IV Cube 4108 MS 08, Hyattsville, MD 20782-2064, USA. Tel.: +1 301 458 4865; lmi3@cdc.gov.

are relied upon to establish population reference ranges, track exposure trends, and prioritize research needs. The NHANES questionnaire also incorporates detailed information about study participants' health insurance, including self-reported Medicaid/CHIP enrollment status. National population health surveys, such as NHANES, are widely used in health services research for policy development and evaluation, as they provide a good source of information on those lacking coverage and thus a good way of assessing the extent to which programs are reaching their target populations [3]. Previous research, however, which has compared Medicaid status reported in surveys with administrative records, has shown that Medicaid enrollment is often underreported on health surveys [4,5]. This phenomenon is referred to as the "Medicaid Undercount". One report using NHANES data which have been linked to the Centers for Medicare and Medicaid Services' Medicaid Analytic eXtract files (CMS MAX) indicates that among 1999–2004 NHANES participants under the age of 18, only 74% of those enrolled in Medicaid actually reported being enrolled (unweighted percentage) [6]. Studies examining the linked 2000–2004 Medicaid Statistical Information System and the 2001–2004 Current Population Survey's (CPS) Child Health Insurance Program data, report that less than two-thirds of those whom the administrative data identify as having Medicaid/CHIP coverage actually report having Medicaid/CHIP coverage in the survey [7,8]. Underestimates of Medicaid participation from research based on survey reports can lead to poor health policy decisions [3].

Using linked files to determine Medicaid and CHIP status may lead to more accurate estimates of program participation and better data for examining the health of program beneficiaries. Linked NHANES-CMS Medicaid files are available through the National Center for Health Statistics' Research Data Center (RDC) for survey years 1999–2004 and Medicaid/CHIP claims files between 1999 and 2009. These files are not public access due to the increased risk of disclosure associated with linked data files. Within the linked dataset, the administrative data provide information regarding monthly enrollment status, eligibility group, and use and costs of services during the coverage period, while survey data capture sociodemographic characteristics, health history (addressed and unaddressed by doctors), dietary habits, health-related behaviors, access to health care, laboratory measures, and physical examination components.

A disadvantage of linked data is that not all survey participants can be linked to administrative files. NHANES participants who do not provide sufficient personal identifiers, such as their social security number or their health insurance claim number are ineligible for linkage. One way to analyze incompletely linked data is to limit analyses to the linkage eligible individuals. However, survey respondents with sufficient personal identification for linkage are self-selected. If the linkage eligible subset differs systematically from those who are not eligible, then eliminating the linkage ineligibles without adjustments could lead to biased estimates.

In a previous project, we compared three methods of addressing the potential bias caused by linkage ineligibility in analyses using the NHANES-CMS Medicaid linked data to examine associations between Medicaid/CHIP status and one health measure in children (serum cotinine levels, a marker of second hand smoke exposure): one that used multiple imputation (MI) [9] to impute the administrative Medicaid/CHIP status of those who are ineligible for linkage, a second that used the linked data restricted to linkage eligible participants with a basic weight adjustment to account for the non-response among linkage ineligibles [10], and a third that used self-reported Medicaid/CHIP status from the survey data. We found that when using the NHANES CMS-MAX linked data, both the MI approach and the weight adjustment approach were appropriate and effective ways to address the biases that result from some survey participants being ineligible for linkage. The survey data alone produced very different estimates, which were presumably biased based on the Medicaid Undercount.

The advantage of the MI approach over the weight adjustment approach is that it incorporates all survey participants into the analysis and is able to include information from a large number of covariates. A disadvantage of the MI approach is that its complexity requires additional statistical expertise and familiarity with restricted-use variables. These variables could vary depending on the analysis, but often include the state, month, and year of the NHANES interview, true variance units (as opposed to publicly released masked variance units) from the NHANES data files, and the Medicaid enrollment period of linked survey participants from the linked administrative data. Since the linked data and other restricted-use variables are accessed through the Research Data Center (RDC) at the National Center for Health Statistics (NCHS), creating MI data prior to conducting a subject-specific analysis requires additional time in the RDC.

Due to the potential complexity of accessing the data and conducting multiple imputation, it is of interest to consider the utility of a general use imputation model. A general use imputation model provides a multi-purpose dataset that is complete with Medicaid/CHIP enrollment status for both linkage-eligible survey participants (from administrative records) and linkage-ineligible survey participants (imputed), which in this case would be used to analyze health measures within the Medicaid/CHIP population or to examine associations between Medicaid/CHIP enrollment and health status. A multi-purpose user dataset makes the MI analysis method accessible to researchers who may not have experience performing multiple imputation themselves or who prefer to conduct multiple studies with the same data files. In a project conducted at the NCHS, a general use imputation model was used to multiply impute missing income data in the National Health Interview Survey (NHIS). The multi-purpose datasets derived from that imputation project continue to be used to inform a wide variety of health analyses [11]. Practically speaking, multi-purpose user datasets increase the efficiency of analysis time and make comparisons across analyses from different researchers easier since the same imputed dataset can be used for multiple analyses.

However, identifying the best general use imputation model is a challenge. While it is well known that including all analysis variables (dependent variables and covariates) in the imputation model is advantageous, it may not always be possible to know in advance all of the analyses that might be performed with a multi-purpose dataset.

The objective of this paper is to compare subject specific imputation models to general use imputation models for a variety of analyses among children in order to assess the effect of using a multi-purpose user dataset with “complete” data on Medicaid/CHIP enrollment status for all NHANES participants (either from the administrative records or imputation) when Medicaid/CHIP status is subject to missingness due to linkage ineligibility. Two general use models were considered. The first only included demographic variables, survey design variables, and predictors of Medicaid/CHIP enrollment. It did not include any potential health variables whose association with Medicaid/CHIP status might be of key interest to researchers. The second included 10 such health variables whose association with Medicaid/CHIP enrollment status might be of interest to researchers.

The motivation behind comparing two general use models was that the first model would provide a valid assessment of whether the multi-purpose dataset would work for analyzing associations between Medicaid/CHIP enrollment and health variables that were not considered when the imputation model was being built, while the second model would provide an example of a more informed imputation model. Comparisons were drawn across the three imputation methods (subject-specific, general use without health variables, general use with 10 health variables) through the analyses of four different health associations: untreated dental caries, obesity, serum cotinine, and attention deficit hyperactivity disorder (ADHD). A preliminary version of this project was published in the 2017 Joint Statistical Meetings Proceedings [12]. While many factors need to be considered when determining whether or how to create a multi-use dataset, including resources and competing priorities, this project helps inform the robustness of such a dataset for multiple analyses.

2. Material and methods

2.1. National Health and Nutrition Examination Survey (NHANES) data

NHANES is a nationally representative survey of the resident, civilian, noninstitutionalized United States population. It is designed to monitor the country's health and nutritional status and includes an interview in the home followed by a standardized physical examination at a specially designed mobile examination center (MEC). Survey participants are selected using a complex, multistage probability sampling design, details of which have been described elsewhere [13]. Sample weights account for oversampling, survey non-response, and post-stratification. During NHANES 1999–2004, oversampled groups included: Mexican-Americans, black persons, low-income persons (at or below 130% of the federal poverty level), and adolescents aged 12–19 years. The oversampling of low-income individuals and adolescents increased the sample size of potential Medicaid/CHIP beneficiaries over what it would have otherwise been had these populations not been oversampled. A proxy provided information for survey participants who were less than 16 years of age and for individuals who could not answer the questions themselves.

The NHANES question on Medicaid/CHIP coverage from 1999–2004 read, "Is the study participant covered by Medicaid/CHIP?" It did not allow for a distinction between the two or for the exclusion of CHIP beneficiaries from analyses. In efforts to be consistent with the survey question, both Medicaid and CHIP were treated as one category in our analyses.

2.2. Centers for Medicare and Medicaid Services' Medicaid Analytic eXtract (CMS MAX) files

Since 1999, Medicaid data have been collected by states and provided to CMS through the Medicaid Statistical Information System (MSIS). These data include enrollee eligibility information, service utilization, and Medicaid claims paid in each quarter of the federal fiscal year. The Medicaid Analytic eXtract (MAX) files are research extracts of MSIS which provide person-level information on demographics, monthly enrollment status, eligibility group, and use and costs of services during the year.

In addition to Medicaid records, the MAX files also contain records from CHIP. CHIP provides health coverage to low-income, uninsured children and pregnant women in families with incomes too high to qualify for state Medicaid programs. CHIP is administered by states according to federal requirements and is funded jointly by the state and federal governments. States may choose whether to provide Medicaid expansion CHIP programs (M-CHIP), which provide the standard Medicaid benefit package to these children, or separate CHIP programs (S-CHIP), which provide coverage that is actuarially equivalent to other health insurance programs, such as those offered to federal and state employees. For the purposes of MSIS, M-CHIP is part of Medicaid, but S-CHIP is not. States are required to report M-CHIP enrollees, but are not required to report S-CHIP enrollees to MSIS. The CMS MAX files include all children enrolled in Medicaid, all children enrolled in M-CHIP, and some children enrolled in S-CHIP. As a result, the combined category used in this study may miss some S-CHIP enrollees [14]. However, as Klerman et al. [7] concluded when analyzing similarly linked CPS-Medicaid/CHIP data, it is a workable solution that helps to mitigate any Medicaid-CHIP confusion since the two are inseparable in the survey for 1999–2004.

2.3. Data linkage

Data linkage between NHANES and the CMS MAX files is performed regularly by the NCHS Data Linkage Program. For NHANES 1999–2004, survey participants were asked to provide their social security number (SSN) and their Medicare health insurance claim number and were informed that by providing this information their survey data would be linked to vital statistics and other records and used for statistical purposes to conduct health-related research. For these years, survey participants were linkage-eligible if they consented to linkage by supplying sufficient personally identifiable information and if their SSN was verified by the Social Security Administration’s Enumeration Verification System [16]. Survey participants were ineligible for linkage if consent was not given or personally identifiable information not provided. Linkage eligible survey participants whose SSN, month and year of birth, and sex exactly match with the CMS MAX files were considered “linked”. This paper uses the most recently linked data available: NHANES 1999–2004 data linked to Medicaid/CHIP claims files between 1999 and 2009. Linked enrollment and claims data for NHANES 2005 through 2012 are expected to be available by the end of 2018. For this analysis the claims data were limited to 1999–2004 to coincide with the survey years.

2.4. Analytic sample

Using NHANES 1999–2004, this study included children ages 2–18 years who participated in the Mobile Examination Center (MEC) exam. Figure 1 indicates how many NHANES 1999–2004 participants aged 2–18 years were linkage eligible, how many of the linkage eligible were linked versus not linked, and how many were ineligible for linkage. For this study, children were identified as linked if they had full or partial enrollment status in

Medicaid, S-CHIP, or M-CHIP within the same state that their interview was conducted for at least one day during the month and year of their interview. Linkage-eligible survey participants who were enrolled outside of that window (different state and/or month and year of interview) were considered not linked. Children who were linkage eligible and were linked to the administrative records were classified as Medicaid/CHIP beneficiaries (n = 3,400), children who were linkage eligible and not linked were classified as non-Medicaid/CHIP beneficiaries (n = 5,923), and children who were ineligible for linkage had unknown Medicaid/CHIP status (n = 2,670). MI was used to impute enrollment status for children who were ineligible for linkage (i.e. had unknown Medicaid/CHIP status). MI was simultaneously used to impute missing information for all other covariates used in the imputation model. Characteristics of the analytic sample overall and by administrative Medicaid/CHIP status are provided in Table 1.

Analytic sample sizes varied depending on which health variable was being analyzed as the dependent variable in the regression analysis. This was because of differences in data collection across the different components of NHANES. Without sample restrictions, data included in the imputation model would have been systematically missing for certain age groups and survey cycles; for example, the early childhood questionnaire, which provided many covariates for the ADHD imputation, was only administered to children under the age of 15 years. Systematic missingness is not appropriate for the traditional MI model. Figure 2 indicates what restrictions were placed on each of the four analyses and the final sample sizes corresponding to each outcome specific analytic sample.

2.5. Multiple imputation

MI was conducted using SAS version 9.3 PROC MI (Fully Conditional Specification option) with 100 imputations. Data were assumed to be missing at random (MAR). Six imputation models were developed: four subject specific imputation models and two general use models. The first general use model included demographic variables, survey design variables, and survey variables related to Medicaid/CHIP enrollment. The second general use model included all of the aforementioned variables, as well as 10 commonly studied health variables. The subject specific models included all of the variables used in the first general use model, as well as the dependent variable of interest, and predictors related to the dependent variable of interest: untreated dental caries, obesity, serum cotinine, or ADHD. Tables 2 and 3 list the variables included in each of the imputation models.

For all imputations, survey design variables included primary sampling units (PSU), typically counties; strata; and sample weights. For technical efficiency, a continuous variable which represents the percentage of Medicaid/CHIP beneficiaries within each PSU based on the linked NHANES CMS MAX data was created to replace the original PSU variable, which had 87 categories. Using PSU level characteristics, rather than PSU indicators, is a technique that was previously implemented when imputing income for the NHIS [11]. In addition to the survey and design variables, a final explanatory variable was created by combining self-reported family income and state of residence with Kaiser's 2004 reports of Medicaid and S-CHIP state income thresholds (http://kff.org/state-category/medicaid-chip/). This variable classifies children as Medicaid eligible, S-CHIP eligible, or neither and was used exclusively as a predictor in the imputation models. It did not serve as a correction to the administrative Medicaid/CHIP or the survey Medicaid/CHIP variables, as there are many ways besides income by which children might be eligible for Medicaid/CHIP.

Linear regression was used to impute continuous variables, logistic regression for binary and ordinal variables, and the discriminant function for all other categorical variables. Citizenship status was an exception; though it was only two categories (citizen, not a citizen) the discriminant function was used. Imputation for all missing variables was performed jointly to fully incorporate the relationship among these variables as well as with aforementioned predictors [15].

Within each analytic sample, sample sizes also varied across imputation methods. All imputation models imputed missing values for all variables that were included in the imputation model. However, variables that were included in the analysis model, but not the imputation model were subject to item non-response and survey participants with item non-response were excluded from analyses. Analyses based on the general use model without health variables had "complete" data for the Medicaid/CHIP enrollment variable and demographic covariates, but in some cases had item nonresponse for dependent variables and/or for covariates that were specific to the dependent variable. Analyses based on the general use model with health variables had "complete" data for the administrative Medicaid/CHIP enrollment variable, the dependent variable, and demographic covariates, but in some cases had item nonresponse for covariates that were specific to the dependent variable of interest. Analyses based on the subject specific imputation models had "complete" data for all variables used in the analyses. The differences in sample size are displayed in the results section (Table 4).

2.6. Analysis

Regression analyses were performed to examine associations between Medicaid/CHIP enrollment and four different health variables: untreated dental caries, obesity, serum cotinine, and ADHD. These health measures were chosen because they are important indicators of children's health, have been previously shown to exhibit differences by socioeconomic status [16–19], and represent a variety of components from the NHANES survey (oral health data, body measurement data, laboratory data, and questionnaire data). Three of the four dependent variables are Leading Health Indicators monitored for Healthy People 2020. Healthy People 2020 strives to increase by 10% the proportion of children and adolescents who used an oral health care system in the last year (from 44.5% to 49%), lower obesity among children by 10% (from 16.1% to 14.5%), and reduce the proportion of children aged 3–11 exposed to secondhand smoke by 10% (from 52.2% to 47%). The fourth dependent variable, ADHD, is the most commonly diagnosed neurobehavioral disorder of childhood [20] and is more prevalent among children with Medicaid than among uninsured or privately insured children [18].

Logistic regression models were fit to examine the association between untreated dental caries and Medicaid/CHIP enrollment, obesity and Medicaid/CHIP enrollment, and ADHD and Medicaid/CHIP enrollment, where complete Medicaid/CHIP enrollment variables were obtained from the MI results above. A log linear model was fit to examine the association between serum cotinine levels and Medicaid/CHIP enrollment. Medicaid/CHIP enrollment was defined as a binary predictor variable: enrolled or not enrolled.

All models controlled for the following sociodemographic characteristics: sex (male/female), race/Hispanic origin (Mexican American, non-Hispanic white, non-Hispanic black, all other races and ethnicities including multi-racial), age at the time of the mobile examination (varied across models: sometimes categorized as 1–5, 6–11, 12–18 and sometimes included as a continuous variable), and ratio of family income to poverty (FIPR, ordinal: $\leqslant 1$, 1.01–2, 2.01–3, 3.01–4, > 4). The untreated dental caries models and the serum cotinine models also controlled for the education of the household reference person ($\leqslant$ High school graduate/GED, some college/associates degree/college graduate or higher). With the exception of sex, all of these variables have been previously shown to be associated with Medicaid/CHIP enrollment [21–23]. In addition, the untreated dental caries models controlled for time since the last dental visit (never, < 6 months, 6–12 months and > 12 months), the serum cotinine models controlled for whether or not someone in the home smokes (yes/no), and the ADHD models controlled for self-reported health status at the time of the household interview (excellent, very good, good, fair, poor).

The FIPR variable is an index for the ratio of self-reported family income and a federal poverty guideline specific to family size, year, and state provided by the Department of Health and Human Services' (HHS) poverty guidelines. The household reference person is the first household member, 18 years of age or older who is listed on the screener questionnaire household member roster who owns or rents the residence where members of the household reside. The education variable for the household reference person is the highest grade or level of education completed by him/her with response categories corresponding to less than 9th grade education, 9–11th grade education (includes 12th grade and no diploma), High school graduate/GED, some college or associates (AA) degree, and college graduate or higher.

All analyses were performed with SAS-callable SUDAAN, version 9.3 PROC REGRESS/PROC RLOGIST, and accounted for the complex survey design. Variance estimates were calculated using the Taylor linearization with replacement method and Student's t-tests were conducted to test the null hypothesis that β coefficients were equal to zero by using a significance level of $p < 0.05$.

For the sake of comparisons, the subject specific imputation model was considered the gold standard. The utility of the two general use imputation models (one without health variables and one with 10 health variables) were assessed by comparing the estimates associated with the general use imputation models to those associated with the subject specific imputation models.

3. Results

Among all linkage eligible survey participants in our sample, 36.5% (unweighted) linked in the same state, month, and year as the NHANES. In other words, among those who were not missing data for administrative Medicaid/CHIP enrollment status, 36.5% (unweighted) were classified as Medicaid/CHIP beneficiaries. Table 5 shows the unweighted percentage of all survey participants in our sample who were either linked or had imputed Medicaid/CHIP enrollment for each of the six different models, as well as the unweighted percentage of linkage ineligible participants with imputed Medicaid/CHIP enrollment for each of the six different models. The (unweighted) percentage of survey participants in our sample who were classified as Medicaid/CHIP beneficiaries based on the imputed datasets was lower for all participants combined then it was among linkage eligible participants (complete cases) and was lower among the linkage ineligible participants (those missing Medicaid/CHIP status) than among the linkage eligible participants (complete cases). Over 100 imputations, the average (unweighted) percent of all survey participants in our sample classified as Medicaid/CHIP beneficiaries ranged from 33.3% (SE = 0.18) using the obesity specific imputation to 36.4% (SE = 0.2) using the cotinine specific imputation. Both general use imputation models classified 34.1% (SE = 0.15 general use without health outcome variables, SE = 0.17 general use with health outcome variables) of survey participants in our sample as Medicaid/CHIP beneficiaries. The average (unweighted) percent of linkage ineligible children classified as Medicaid/CHIP beneficiaries ranged from 25.4% (SE = 0.81) using the obesity specific imputation to 28.1% (SE = 0.90) using the cotinine specific imputation.

Table 4 presents the results for the regression analyses and Fig. 3 shows the exponentiated beta coefficient and 95% confidence interval corresponding to Medicaid/CHIP enrollment within each regression.

Among all children, aged 2 to 18 years, who completed both the oral health exam and the dietary recall, 22.6% (SE = 0.92) had untreated dental caries at the time of the exam. This prevalence was 31.8% (0.02) among linked Medicaid/CHIP beneficiaries, 19.6% (0.01) among linkage eligible non-Medicaid/CHIP beneficiaries, and 21.4% (0.01) among linkage ineligibles. In the untreated dental caries regression model, the odds ratio corresponding to Medicaid/CHIP enrollment was 1.10 [95% CI: (0.86, 1.42)] using the subject specific imputation. Comparatively, the odds ratio corresponding to the general use imputation without health outcomes was 1.13 [95% CI: (0.88, 1.46)] and the odds ratio corresponding to the general use imputation with health outcomes was 1.12 [95% CI: (0.87, 1.44)]. In terms of inference, all three imputation methods indicate that there is no statistically significant association between Medicaid/CHIP enrollment and whether or not a child has untreated dental caries.

Among all children, aged 3–18 years, who participated in the MEC examination, 15.5% (SE = 0.54) were obese at the time of the exam. This prevalence was 18.3% (SE = 1.1) among linked Medicaid/CHIP beneficiaries, 15% (SE = 0.7) among linkage eligible non-Medicaid/CHIP beneficiaries, and 13.9% (SE = 0.8) among linkage ineligibles. In the obesity regression model, the odds ratio corresponding to Medicaid/CHIP enrollment was 1.13 [95% CI: (0.93, 1.38)] using the subject specific imputation model. Comparatively, the odds ratios corresponding to the general use imputation models were both 1.17 [95% CI: (0.96, 1.43)]. In terms of inference, all three imputation methods indicate that there is no statistically significant association between Medicaid/CHIP enrollment and obesity.

Among all non-smoking children, aged 3–15 years, who participated in the MEC examination, the average serum cotinine level was 0.60 ng/mL (SE = 0.05) and the geometric mean was 0.13 ng/mL (SE = 0.01). These values were 0.97 ng/mL (SE = 0.08) and 0.32 ng/mL (SE = 0.04), respectively, among the linked Medicaid/CHIP beneficiaries, 0.47 ng/mL (SE = 0.04) and 0.09 ng/mL (SE = 0.008), respectively, among the linkage eligible non-Medicaid/CHIP beneficiaries, and 0.48 ng/mL (SE = 0.07) and 0.11 ng/mL (SE = 0.01), respectively, among linkage ineligibles. In the serum cotinine regression model, the exponentiated beta coefficient corresponding to Medicaid/CHIP enrollment was 1.41 [95% CI: (1.20, 1.65)] using the subject specific imputation. This corresponds to a 41% increase in average serum cotinine levels among Medicaid/CHIP beneficiaries as compared to non-Medicaid/CHIP beneficiaries. Comparatively, the exponentiated beta coefficient was 1.35 [95% CI: (1.15, 1.58)] using the general use imputation without health outcomes and 1.38 [95% CI: (1.19, 1.62)] using the general use imputation with health outcomes. These correspond to increases of 35% and 38%, respectively in the average serum cotinine levels of Medicaid/CHIP beneficiaries as compared to non-Medicaid/CHIP beneficiaries. Thus, all three imputation methods indicate that after controlling for sex, race, age, FIPR, the education of the household reference person, and whether or not there is a smoker in the home, the average serum cotinine levels of Medicaid/CHIP beneficiaries was higher than that of non-Medicaid/CHIP beneficiaries.

Among all children, aged 6 to 15 years, who participated in the MEC examination, 9.3% (SE = 0.54) identified as having ever been told by a doctor or health professional that they have ADHD. This prevalence was 12.9% (SE = 1.3) among linked Medicaid/CHIP beneficiaries, 7.8% (SE = 0.7) among linkage eligible non-Medicaid/CHIP beneficiaries, and 9.2% (SE = 1.4) among linkage ineligibles. In the ADHD regression models, the odds ratio corresponding to Medicaid/CHIP enrollment was 2.08 [95% CI: (1.43, 3.05)] using the subject specific imputation, 1.83 [95% CI: (1.26, 2.66)] using the general use imputation without health outcome variables, and 1.91 [95% CI: (1.31, 2.78)] using the general use imputation with health outcome variables. In terms of inference, all three methods indicate that after controlling for sex, age, race, FIPR, and self-reported health status, the odds of children having ADHD for those enrolled in Medicaid/CHIP is 1.83–2.08 times higher than that of children who are not enrolled. The association was strongest using the subject specific imputation model and the general use model with health outcomes presented an estimate that was closer to estimate produced by the subject specific imputation model than the general use model without health outcomes.

Table 6 presents the relative differences between the exponentiated β coefficients (in most cases odds ratios) associated with the subject specific imputation models and the exponentiated β coefficients associated with the two different general use imputation models. The odds ratio corresponding to the general use imputation without health outcomes was within 12% of the odds ratio corresponding to the subject specific imputation in the ADHD analysis, 4% in the obesity analysis, and 3% in the untreated dental caries analysis. Similarly, for the log linear model, the estimated exponentiated β coefficient corresponding to the general use imputation without health outcomes was within 4% of the estimate corresponding to the subject specific imputation. The estimated percent increase in average serum cotinine levels among Medicaid/CHIP beneficiaries corresponding to the general use imputation without health outcomes was within 15% of the estimate corresponding to the subject specific imputation.

The odds ratio corresponding to the general use imputation with health outcomes was within 8% of the odds ratio corresponding to the subject specific imputation in the ADHD analysis, 4% in the obesity analysis, and 2% in the untreated dental caries analysis. For the log-linear model, the estimated exponentiated β coefficient corresponding to the general use imputation with health outcomes was within 2% of the estimate corresponding to the subject specific imputation, which corresponded to a 7% change in the estimate of percent increase in average serum cotinine levels among Medicaid/CHIP beneficiaries.

4. Discussion

The results demonstrate that in the cases examined here, general use imputation models provide estimates for the effect of Medicaid/CHIP coverage that are comparable to the estimates produced by subject specific imputation models. Assuming the subject specific imputation models were ‘gold’ standards, using a general use imputation model with 10 commonly analyzed health variables led to relative differences from 2% to 8%. Compared to the gold standard, using a general use imputation model without health variables, relative differences ranged from 3% to 15%.

Exact cut points for determining meaningful differences across models are somewhat arbitrary and depend on the specific study and outcome of interest. Statistical testing was not done since the same survey participants were included in each model and the models therefore lacked independence. In a discussion by Rothman et al. on variable selection for epidemiological studies in the context of confounding, a change in the effect of interest (relative risk) of greater than 10% after stratification on a variable indicates that the adjustment should be retained [24]. Although not directly applicable to this study, the Rothman guideline is one way to inform comparative judgements of effect sizes across imputation models. Compared to the subject specific imputation models, the general use imputation model with health measure variables produced relative differences in odds ratio estimates and estimates of percent change that were all within the 10% threshold. Compared to the subject specific imputation models, the general use imputation model without health measure variables produced some relative differences that were slightly higher than the 10% threshold.

Across all four dependent variable analyses, both of the general use imputation models offered similar inferences to that of the subject specific imputation model. After controlling for relevant covariates, all three imputation methods led to the conclusions that for children enrolled in Medicaid/CHIP the odds of having ADHD is higher than that of children who are not enrolled and that the average serum cotinine levels of Medicaid/CHIP beneficiaries is about one-third higher than that of non-Medicaid/CHIP beneficiaries. All three imputation methods also led to the conclusions that there are not statistically significant associations between Medicaid/CHIP enrollment and obesity or between Medicaid/CHIP enrollment and untreated dental caries.

For this imputation project we were most interested in examining how the imputed Medicaid/CHIP enrollment status variable performed when examining associations between Medicaid/CHIP enrollment and health status for children. We were not trying to accurately predict Medicaid/CHIP enrollment at the person-level. However, the distribution of Medicaid/CHIP beneficiaries after imputation helped to confirm that all 6 imputation models performed similarly. The estimated percentages of all children enrolled in Medicaid/CHIP based on the imputation models were all within 10% of one another, while the estimated percentages of linkage ineligible children enrolled in Medicaid/CHIP based on the imputation models were all within 11% of one another. Though the estimated percentage of Medicaid/CHIP beneficiaries was lower among those who were ineligible for linkage as compared to those who were linkage eligible, this followed the expected pattern given that higher income households and households with higher education levels are less likely to provide personally identifiable information [14].

Based on these examples, we make two observations regarding a general use imputation model relative to a subject-specific model (our gold standard). First, the effectiveness of the general use imputation models demonstrate that even when there are relative differences of up to 12% across beta coefficients, overall conclusions regarding associations between Medicaid/CHIP enrollment and health status are not highly affected by the use of general use imputation models as compared to subject specific imputation models. Second, estimates of the effect of Medicaid/CHIP enrollment on health status produced from the general use imputation models with health variables were closer to the estimates produced from subject specific imputation models than estimates produced from general use imputation models without health variables.

There are two practical advantages of a multi-purpose user dataset. First, it facilitates analyses for researchers conducting more than one Medicaid/CHIP-related project, as they can use a common multiply imputed dataset for each study. Second, it could allow for consistent comparisons across analyses using the MI method to account for the potential bias due to linkage ineligibles. In some cases, however, a subject specific imputation model may still be preferred in order to maximize the number of survey participants included in the final analysis, as using a multi-purpose dataset may exclude children who have item non-response for any dependent variables or covariates not originally included in the general use imputation model.

These analyses did not compare the MI approach with other methods of adjusting for linkage ineligibility. It is not clear whether using MI to account for Medicaid/CHIP enrollment status among those who are linkage-ineligible is comparable to using the currently recommended weight adjustment approach [10]. Previous work demonstrated that when examining the association between Medicaid/CHIP and serum cotinine, the subject specific imputation analysis was comparable to the weight adjustment approach. However, the weight adjustment approach has not been directly compared to the general use imputation analyses, nor have comparisons between the subject specific MI approach and the weight adjustment approach been extended to other health measures besides serum cotinine.

Moreover, while in theory, subject specific imputation models, which include all the variables that will be used in the analyses, are superior to general use imputation models [25], the subject specific imputation models presented here are models developed to the best of our ability with the data that was available. As such, they are dependent upon the availability of necessary covariates. While they function well as analytic tools for comparison and represent the subject specific imputation models that would most likely be used in an analysis of these health outcomes, they are not true gold standards.

Finally, though efforts were made to choose a variety of dependent variables from different components of the NHANES (oral health data, body measurement data, laboratory data, and questionnaire data), we do not know if studies of other health variables would lead to similar results. General-use imputation models may not perform as well in relation to subject specific imputation models for all dependent variables within the NHANES. Likewise, general use models may not perform as well among adult populations as they do for children or for imputation models that differentiate between Medicaid and CHIP status.

5. Conclusion

This study illustrates that for four selected health measures, untreated dental caries, obesity, serum cotinine, and ADHD, using a general use imputation model to produce a multi-purpose user dataset with “complete” Medicaid/CHIP enrollment status for survey participants (either from the linked data or from imputation) is an alternative to subject specific models for performing analyses of association between Medicaid/CHIP enrollment and health status when linked data are subject to missingness for administrative Medicaid/CHIP status due to linkage eligibility. The results suggest that including a variety of potential dependent variables might improve the imputation, but that analyses need not be limited to using health variables that are included in the imputation model. However, further exploration of general use imputation models could provide additional insights, as the best general use imputation model for this task is unknown and there may be better methods for addressing the potential biases targeted here.

Acknowledgments

The authors gratefully acknowledge subject level experts at NCHS for their guidance and review. These include Lisa Mirel and Cordell Golden (linked data), Debra Brody (cotinine), Eleanor Fleming (oral health), and Lara Akinbami (ADHD).

Disclaimer: The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the National Center for Health Statistics, Centers for Disease Control and Prevention.

References

[1]. The Kaiser Commission on Medicaid and the Uninsured. Medicaid A Primer: Key Information on the Nation’s Health Coverage Program for Low-Income People. The Henry J. Kaiser Family Foundation Report number: 7334–05, 2013 Available from: https://kaiserfamilyfoundation.files.wordpress.com/2010/06/7334-05.pdf [accessed on September 1, 2017].

[2]. Truffer CJ, Wolfe CJ, Rennie KE. 2015 Actuarial Report on the Financial Outlook for Medicaid. Office of the Actuary, Centers for Medicare & Medicaid Services, United States Department of Health & Human Services, 2015 Available from: https://www.medicaid.gov/medicaid/financing-

and-reimbursement/downloads/medicaid-actuarial-report-2015.pdf [accessed on September 1, 2017].

[3]. Call KT, Davidson G, Davern M, Nyman R. Medicaid Undercount and Bias to Estimates of Uninsurance: New Estimates and Existing Evidence. Health Services Research. 2008; 43(3): 901–914. doi: 10.1111/j.1475-6773.2007.00808.x. [PubMed: 18546545]

[4]. Davern Michael. Research Project to Understand the Medicaid Undercount: 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 U.S. CensusBureau. 2007 Available from: https://www.census.gov/did/www/snacc/docs/SNACC_Phase_I_Full_Report.pdf [accessed on September 1, 2017].

[5]. Davern M, Klerman JA, Baugh DK, Call KT, Greenberg GD. An Examination of the Medicaid Undercount in the Current Population Survey: Preliminary Results from Record Linking. Health Services Research. 2009; 44(3): 965–87. doi: 10.1111/j.1475-6773.2008.00941.x. [PubMed: 19187185]

[6]. Mirel LB, Simon AE, Golden C, Duran CR, Schoendorf KC. Concordance between survey report of Medicaid enrollment and linked Medicaid administrative records in two national studies National Center for Health Statistics. National Health Statistics Reports number: 72, 2014.

[7]. Klerman J, Plotzke MR, Davern M. CHIP Reporting in the CPS. Medicare & Medicaid Research Review. 2012; 2(3). doi: 10.5600/mmrr.002.03.b01.

[8]. Davern M, Klerman JA, Baugh DK, Call KT, Greenberg GD. An Examination of the Medicaid Undercount in the Current Population Survey: Preliminary Results from Record Linking. Health Services Research. 2009; 44(23): 965–87. doi: 10.1111/j.1475-6773.2008.00941.x. [PubMed: 19187185]

[9]. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley; 1987.

[10]. Judson DH, Parker JD, Larsen MD. Adjusting sample weights for linkage-eligibility using SUDAAN. National Center for Health Statistics 2013 Available from: http://www.cdc.gov/nchs/data/datalinkage/adjusting_sample_weights_for_linkage_eligibility_using_sudaan.pdf [accessed on September 1, 2017].

[11]. Schenker N, Raghunathan TE, Chiu P, Makuc DM, Zhang G, Cohen AJ. Multiple Imputation of Missing Income Data in the National Health Interview Survey. Journal of the American Statistical Association. 2006; 101(475): 924–933. Available from: doi: 10.1198/016214505000001375.

[12]. Rammon J, He Y, Parker JD. Multiple Imputation to Enhance the CMS-Medicaid Linked Data. Papers and Proceedings, Joint Statistical Meetings, 2018.

[13]. Curtin LR, Mohadjer L, Dohrmann S, Montaquila JM, Kruszon-Moran D, Mirel LB, Carroll MD, Hirsch R, Schober S, Johnson CL. The National Health and Nutrition Examination Survey: Sample design, 1999–2006. National Center for Health Statistics. Vital Health Statistics Report: 2(155). 2012.

[14]. Golden C, Driscoll AK, Simon AE, Judson DH, Miller EA, Parker JD. Linkage of NCHS population health surveys to administrative records from Social Security Administration and Centers for Medicare & Medicaid Services. National Center for Health Statistics. Vital Health Statistics Report: 1(58); 2015.

[15]. Collins LM, Schafer JL, Kam C. A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychological Methods. 2001; 6(4): 330–351. [PubMed: 11778676]

[16]. Bagramian RA, Garcia-Godoy F, Volpe AR. The global increase in dental caries. A pending public health crisis. American Journal of Dentistry. 2009; 21(1).

[17]. Environmental Protection Agency. America's Children and the Environment, Biomonitoring: Cotinine. 2016 Available from: https://www.epa.gov/ace/ace-biomonitoring-cotinine [accessed 1 September 2017].

[18]. Pastor PN, Reuben CA. Diagnosed attention deficit hyperactivity disorder and learning disability: United States, 2004–2006. Vital Health Statistics. 2008; 10(237).

[19]. Wang Y, Zhang Q. Are American children and adolescents of low socioeconomic status at increased risk of obesity? Changes in the association between overweight and family income

between 1971 and 2002. The American Journal of Clinical Nutrition. 2006; 84(4): 707–716. 10.1093/ajcn/84.4.707. [PubMed: 17023695]

[20]. Visser SN, Bitsko RH, Danielson ML, Perou R, Blumberg SJ. Increasing Prevalence of Parent-reported Attention-deficit/Hyperactivity Disorder Among Children. Morbidity & Mortality Weekly Report. 2010; 59(44): 1439–1443. [PubMed: 21063274]

[21]. Dubay LC, Kenney GM. The Effects of Medicaid Expansions on Insurance Coverage of Children. The Future of Children. 1996; 6(1): 152–161. Available from: doi: 10.2307/1602499. [PubMed: 8689258]

[22]. Kincheloe J, Frates J, Brown ER. Determinants of children's participation in California's Medicaid and SCHIP programs. Health Services Research. 2007; 42(2): 847–66. Available from: doi: 10.1111/j.1475-6773.2006.00624.x. [PubMed: 17362221]

[23]. Simon AE, Driscoll A, Gorina Y, Parker JD, Schoendorf KC. A Longitudinal View of Child Enrollment in Medicaid. Pediatrics. 2013; 132(656). Available from: doi 10.1542/peds.2013-1544.

[24]. Rothman KJ, Sander G, Lash TL. Modern Epidemiology. Third Edition. Philadelphia: Lippincot Williams & Wilkins; 2008.

[25]. Van Buuren S Flexible Imputation for Missing Data. Chapman and Hall/CRC; 2012.

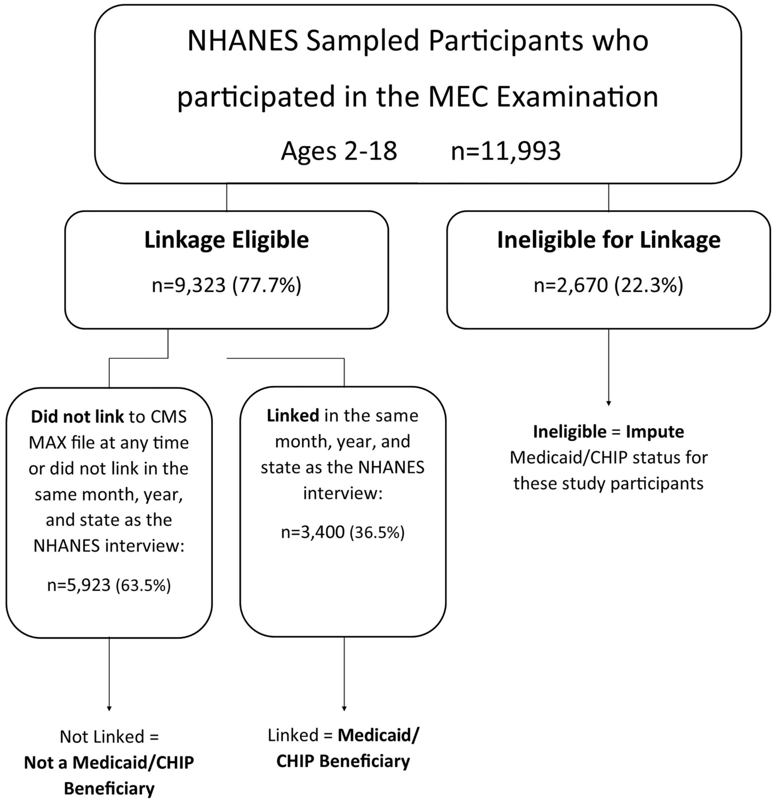

Fig. 1.

Medicaid/CHIP classification based on administrative data. National Health and Nutrition Examination Survey data linked to Centers for Medicare and Medicaid Services' Medicaid data:1999–2004.

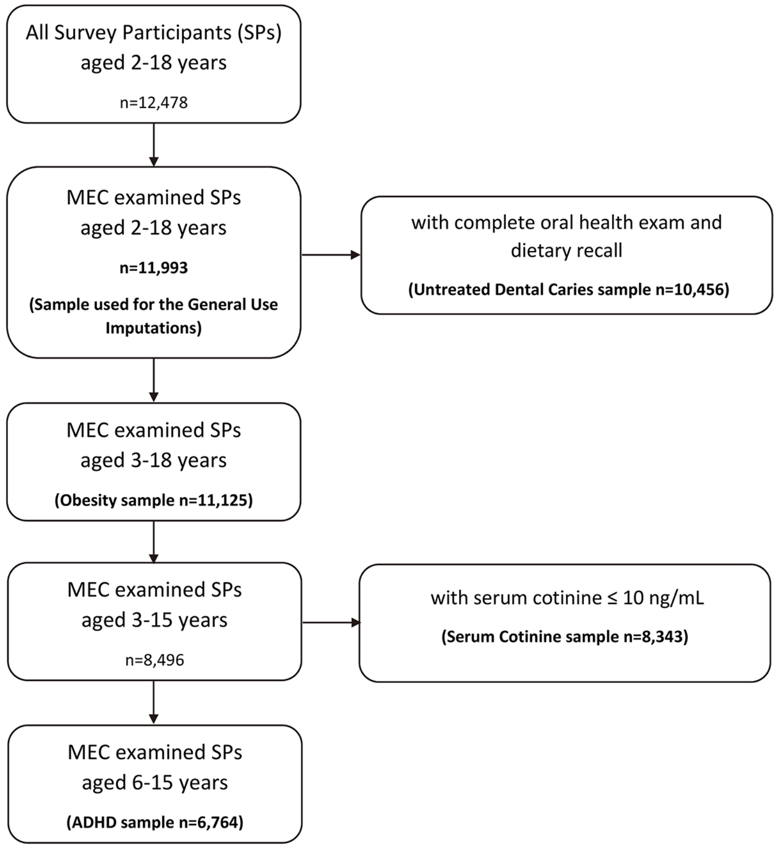

Fig. 2.

Understanding the sample sizes. National Health and Nutrition Examination Survey linked to Centers for Medicare and Medicaid Services' Medicaid data: 1999–2004.

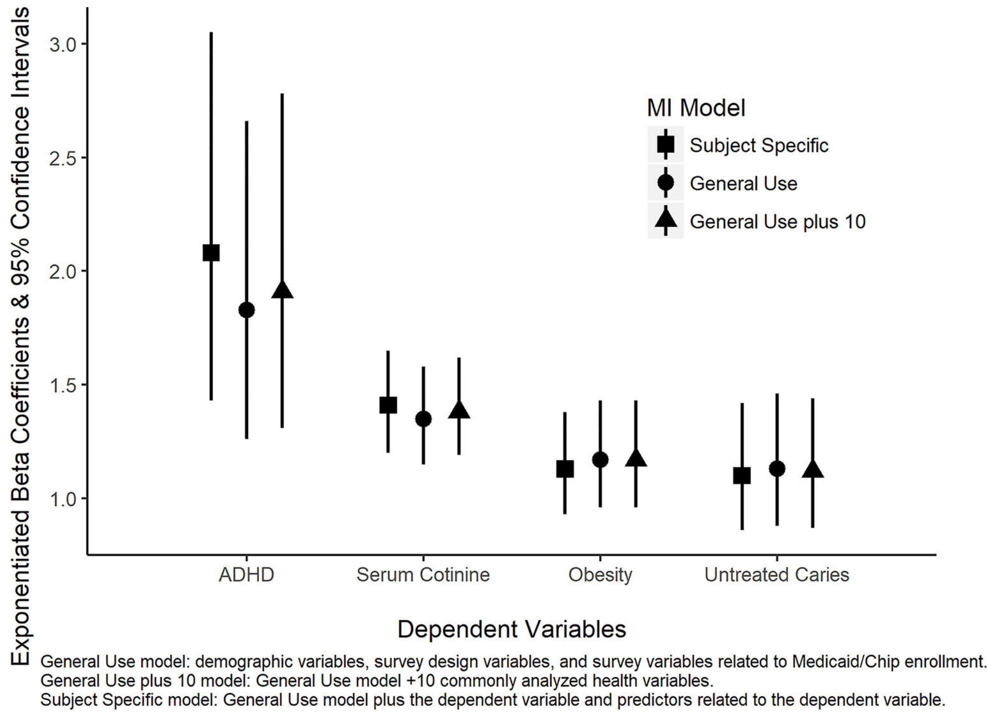

Fig. 3.

Comparison of exponentiated beta coefficients across imputation models. National Health and Nutrition Examination Survey linked to Centers for Medicare and Medicaid Services' Medicaid data: 1999–2004.

Table 1

Characteristics of the Analytic Sample by Administrative Medicaid/CHIP Status: National Health and Nutrition Examination Survey (1999–2004) linked to Centers for Medicare and Medicaid Services Medicaid Analytic eXtract files (1999–2009), sampled children aged 2–18 years who participated in the MEC examination

	Overall	Eligible for linkage		Ineligible for linkage
		Enrolled in Medicaid/CHIP *"linked"*	**Not enrolled in Medicaid/CHIP *"not linked"***	**Medicaid/CHIP enrollment unknown**
n	11,993	3,400	5,923	2,670
Gender[2]				
Male	0.51	0.53	0.52	0.45
Female	0.49	0.47	0.48	0.55
Age category[1,2]				
2–5 years	0.23	0.30	0.22	0.17
6–11 years	0.36	0.39	0.35	0.33
12–18 years	0.42	0.31	0.43	0.50
Race/ethnicity[1,2]				
Mexican American	0.12	0.13	0.11	0.15
Non-Hispanic White	0.60	0.44	0.68	0.55
Non-Hispanic Black	0.15	0.27	0.09	0.16
All other ethnicities including multi-racial	0.13	0.15	0.12	0.14
Poverty income ratio[1,2]				
⩽ 100%	0.22	0.54	0.11	0.19
100% < PIR ⩽ 200%	0.23	0.29	0.21	0.21
200% < PIR ⩽ 300%	0.16	0.06	0.20	0.14
300% < PIR ⩽ 400%	0.12	0.02	0.16	0.12
400% < PIR ⩽ 500%	0.21	0.01	0.29	0.19
Non-response	0.07	0.08	0.04	0.14
Survey Medicaid/CHIP response[1,2]				
Medicaid/CHIP	0.21	0.68	0.05	0.13
Other insurance	0.66	0.23	0.82	0.68

	Overall	Eligible for linkage		Ineligible for linkage
		Enrolled in Medicaid/CHIP *"linked"*	**Not enrolled in Medicaid/CHIP *"not linked"***	**Medicaid/CHIP enrollment unknown**
Uninsured	0.12	0.06	0.13	0.18
Non-response	0.01	0.03	0.008	0.02
Untreated dental caries[1]				
Aged 2–18 years with complete oral health exam and dietary recall				
Present	0.23	0.32	0.20	0.21
Not present	0.77	0.68	0.80	0.79
Obesity[1]				
Aged 3–18 years				
Currently obese	0.16	0.18	0.15	0.14
Not currently obese	0.82	0.79	0.83	0.84
Non-response	0.02	0.03	0.02	0.02
Serum cotinine (ng/mL)[1,2]				
Non-smokers, aged 3–15 years				
Mean	0.60 (0.05)	0.97 (0.08)	0.47 (0.04)	0.48 (0.07)
Geometric mean	0.13 (0.11)	0.32 (0.04)	0.09 (0.008)	0.11 (0.01)
Attention deficit hyperactivity disorder[1]				
Aged 6–15 years				
Ever diagnosed	0.09	0.13	0.08	0.09
Never diagnosed	0.91	0.87	0.92	0.91
Non-response	0.002	0.003	0.001	0.003

[1] Statistically significant association with Medicaid/CHIP enrollment at the $\alpha = 0.05$ level (among those who are linkage eligible).

[2] Statistically significant association with linkage eligibility at the $\alpha = 0.05$ level.

Note: Categorical variables: weighted proportions. Note: Continuous variables: weighted means (standard error). Note: Medicaid/CHIP enrollment defined as enrollment for at least one day during the month and year of the NHANES interview. Note: Poverty Index Ratio is an index for the ratio of self-reported family income divided by the federal poverty guideline specific to family size, year, and state.

Table 2

Variables used in the two general use models. National Health and Nutrition Examination Survey (1999–2004) linked to Centers for Medicare and Medicaid Services' Medicaid data (1999–2009)

General use model without health measures and all other imputations	General use model with health measures
Gender	*All variables used in the general use model without health measures, plus...*
Age	
Race/ethnicity	Untreated dental caries (binary)
Ratio of family income to poverty	ADHD (binary)
Nativity	Asthma (binary)
Citizenship status	BMI category (underweight, normal weight, overweight, obese)
Age of household reference person	
Education of household reference person	Serum cotinine
Nativity of household reference person	Blood lead
Self-reported general health status	Hemoglobin
Average # of health care visits each year	Total cholesterol
Does SP have a routine place for health care	C-Reactive protein
Home ownership (binary)	Vitamin B12
What type of home SP lives in	
Census region	
Medicaid/CHIP enrollment based on self-report	
Medicaid/CHIP eligibility status based on self-reported income	
Average administrative Medicaid/CHIP enrollment across the primary sampling unit	
Strata	
Exam weights	
Interview weights	
Dietary Interview weights	

Table 3

Variables used in the untreated dental caries, obesity, serum cotinine, and attention deficit hyperactivity disorder subject specific models. National Health and Nutrition Examination Survey (1999–2004) linked to Centers for Medicare and Medicaid Services' Medicaid data (1999–2009)

Untreated dental caries	Obesity	Serum cotinine	Attention deficit hyperactivity disorder (ADHD)
All variables used in the general use model without health measures, plus...	***All variables used in the general use model without health measures, plus...***	***All variables used in the general use model without health measures, plus...***	***All variables used in the general use model without health measures, plus...***
Untreated dental caries (binary)	BMI category (underweight, normal weight, overweight, obese)	Serum cotinine	ADHD (binary)
WIC status (Special Supplemental Nutrition Program for Women, Infants, and Children)	Body weight	Mother smoked while pregnant	Seen a mental health professional in the last 12 months
		# of cigarettes/day smoked in the home	Mother smoked while pregnant
	Standing height		Smoker in the home now
Dental sealants	Waist circumference		Currently taking medications that are typically prescribed for ADHD
Time since last dental visit	Triceps skin fold		
Reason for last dental visit	Sub scapular skin fold		
Categorical BMI	Hemoglobin		Birthweight
Total number of carbs eaten yesterday (gm)	Total cholesterol		Blood lead levels
	C-Reactive Protein (CRP)		Maternal age at birth
Total plain water drank yesterday (gm)	# times/week eat restaurant food		Categorical BMI
Candy eaten yesterday (# of times)	# hours/day spend time watching TV, playing video games, or on the computer		# of hours/day spend time watching TV, playing video games, or on the computer
Soda drank yesterday (# of times)			

Table 4

Beta coefficients and odds ratios associated with Medicaid/CHIP enrollment status (enrolled compared to not enrolled). National Health and Nutrition Examination Survey linked to Centers for Medicare and Medicaid Services' Medicaid data: 1999–2004

Outcome	Subject specific MI			General use MI[6] no health outcomes			General use MI[7] 10 health outcomes		
	β (SE)	*p*	*n*	*β* (SE)	*p*	*n*	*β* (SE)	*p*	*n*
Adjusted[1] untreated dental	0.10 (0.13)	0.44	10,456	0.13 (0.13)	0.33	9,818	0.11 (0.13)	0.38	9,818
Caries model[2]	*1.10 (0.86, 1.42)*			*1.13 (0.88, 1.46)*			*1.12 (0.87, 1.44)*		
Odds ratio (95% CI)									
Adjusted[1] obesity model[3]	0.13 (0.10)	0.2	11,125	0.16 (0.10)	0.11	10,900	0.16 (0.10)	0.11	11,125
Odds ratio (95% CI)	*1.13 (0.93, 1.38)*			*1.17 (0.96, 1.43)*			*1.17 (0.96, 1.43)*		
Adjusted[1] cotinine model[4]	0.34 (0.08)	0.0002	8,343	0.30 (0.08)	0.0005	6,739	0.32 (0.08)	0.0001	8,238
Exponentiated β (95% CI)	*1.41 (1.20, 1.65)*			*1.35 (1.15, 1.58)*			*1.38 (1.19, 1.62)*		
Adjusted[1] ADHD model[5]	0.73 (0.19)	0.0003	6,764	0.61 (0.18)	0.002	6,750	0.65 (0.19)	0.0012	6,764
Odds ratio (95% CI)	*2.08 (1.43, 3.05)*			*1.83 (1.26, 2.66)*			*1.91 (1.31, 2.78)*		

[1] All adjusted models control for sex, race/Hispanic origin, age, and poverty-income ratio.

[2] Children aged 2–18 years at the time of the MEC examination with a complete oral health exam and complete 24-hr dietary recall; additional covariates included education of household reference person and time since the last dental visit.

[3] Children aged 3–18 years at the time of the MEC examination.

[4] Children aged 3–15 years at the time of the MEC examination with serum cotinine levels $\leqslant$ 10 ng/mL; additional covariates included education of household reference person and whether or not there is a smoker in the home.

[5] Children aged 6–15 years at the time of the MEC examination; additional covariates included self-reported health status.

[6] Children with item-nonresponse for the dependent variable or for specific covariates related to the dependent variable not included in the analyses.

[7] Children with item non-response for specific covariates related to the dependent variable not included in the analyses.

Table 5

Distribution of Medicaid/CHIP beneficiaries. Comparing the distribution of Medicaid/CHIP beneficiaries among linkage eligible children to the distribution after imputation using six different imputation models. National Health and Nutrition Examination Survey data linked to Centers for Medicare and Medicaid Services' Medicaid data: 1999–2004

	Unweighted percentage of all children identified as Medicaid/CHIP beneficiaries after imputation[1]		Unweighted percentage of children ineligible for linkage identified as Medicaid/CHIP beneficiaries after imputation[2]	
	Average[3]	**Standard error**	**Average[3]**	**Standard error**
Linkage eligible survey participants[4]	36.5		36.5	
Imputation model				
General use	34.1	0.15	26.0	0.68
General use + 10 health outcome variables	34.1	0.17	25.8	0.76
Attention deficit hyperactivity disorder (ADHD)	34.5	0.22	26.4	0.93
Serum cotinine	36.4	0.20	28.1	0.90
Obesity	33.3	0.18	25.4	0.81
Untreated dental caries	33.2	0.17	25.4	0.77

[1] Linkage eligible children who were linked in the same state, month, and year as the NHANES and linkage ineligible children with imputed Medicaid/CHIP enrollment.

[2] Children ineligible for linkage with imputed Medicaid/CHIP enrollment.

[3] Average across 100 imputations.

[4] Complete Cases.

Table 6

Relative difference in exponentiated β coefficients across imputation models. National Health and Nutrition Examination Survey linked to Centers for Medicare and Medicaid Services' Medicaid data: 1999–2004

	Subject specific imputation model "Gold Standard"	**General use imputation model without health measure variables**	**General use imputation model with 10 health measure variables**
	Estimate and confidence interval	***Percent change*** [1]	***Percent change*** [1]
Untreated dental caries (odds ratio)	1.10 (0.86, 1.42)	3%	2%
Obesity (odds ratio)	1.13 (0.93, 1.38)	4%	4%
Serum cotinine (exponentiated β)	1.41 (1.20, 1.65)	4%	2%
corresponding average increase among Medicaid beneficiaries	41%	15%	7%
ADHD (odds ratio)	2.08 (1.43, 3.05)	12%	8%

NOTE: All odds ratios and exponentiated β coefficients compare Medicaid/CHIP beneficiaries with non-beneficiaries.

[1] $Percent\ change = \dfrac{\text{“}Gold\ Standard\text{”}\ Odds\ Ratio/Estimate\text{-}General\ Use\ Odds\ Ratio/Estimate}{\text{“}Gold\ Standard\text{”}\ Odds\ Ratio/Estimate}$.