Published in final edited form as:
Health Serv Outcomes Res Methodol. 2018 August 16; 19(2-3): 87–105. doi:10.1007/s10742-018-0186-3.

Accounting for study participants who are ineligible for linkage: a multiple imputation approach to analyzing the linked National Health and Nutrition Examination Survey and Centers for Medicare and Medicaid Services' Medicaid data

Jennifer Rammon[1], **Yulei He**[1], **Jennifer D. Parker**[1]

[1]Division of Research and Methodology, National Center for Health Statistics, Centers for Disease Control and Prevention, 3311 Toledo Road, Mailstop P-08, Hyattsville, MD 20782-2064, USA

Abstract

Data from the National Health and Nutrition Examination Survey have been linked to the Center for Medicare and Medicaid Services' Medicaid Enrollment and Claims Files for the survey years 1999–2004. The linked data are produced by the National Center for Health Statistics' (NCHS) Data Linkage Program and are available in the NCHS Research Data Center. This project compares the usefulness of multiple imputation to account for data linkage ineligibility and other survey nonresponse with currently recommended weight adjustment procedures. Estimated differences in environmental smoke exposure across Medicaid/Children's Health Insurance Program (CHIP) enrollment status among children ages 3–15 years are examined as a motivating example. Comparisons are drawn across the three different estimates: one that uses MI to impute the administrative Medicaid/CHIP status of those who are ineligible for linkage, a second that uses the linked data restricted to linkage eligible participants with a basic weight adjustment, and a third that uses self-reported Medicaid/CHIP status from the survey data. The results indicate that estimates from the multiple imputation analysis were comparable to those found when using weight adjustment procedures and had the added benefit of incorporating all survey participants (linkage eligible and linkage ineligible) into the analysis. We conclude that both multiple imputation and weight adjustment procedures can effectively account for survey participants who are ineligible for linkage.

Keywords

Children; CHIP; Multiple imputation; NHANES Medicaid linked data; Serum cotinine

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 (The Kaiser Commission on Medicaid and the Uninsured 2013). In

Jennifer Rammon, lmi3@cdc.gov.

2014, children represented 43% of overall Medicaid enrollment and 17% of all Medicaid expenditures (Truffer et al. 2015). 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 both in-house interviews and physical examinations. The NHANES biomarkers 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. Surveys, such as NHANES, are widely used in health services research for policy development and evaluation, as they provide the only source of information on those lacking coverage and thus the only means of assessing the extent to which programs are reaching their target populations (Call et al. 2008). 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 (Davern 2007; Davern et al. 2009). 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/CHIP actually reported being enrolled (unweighted percentage, Mirel et al. 2014). Similar analyses with other surveys, such as the Current Population Survey's (CPS) Child Health Insurance Program data, report that less than two-thirds of those whom the administrative data implies having Medicaid/CHIP actually report having Medicaid/CHIP in the survey (Klerman et al. 2012; Davern et al. 2009). Underestimates of Medicaid participation from research based on survey reports can lead to poor health policy decisions (Call et al. 2008).

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. Within the linked dataset, the administrative data provides information regarding monthly enrollment status, eligibility group, and use and costs of services during the year, while survey data captures 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.

It is of interest, therefore, to develop effective methods for handling the potential bias due to linkage ineligibles. This can be done by treating the outcomes from the cases who refuse linkage as missing data (nonresponse) in the linked dataset and applying statistical methods that have been developed to handle missing data. NCHS currently recommends adjusting the sample weights to account for the missingness that results from survey participants who are ineligible for linkage (Judson et al. 2013). An alternative approach is to multiply impute the Medicaid/CHIP status of survey participants who are ineligible for linkage (Little and Rubin 2002).

In a recent example, multiple imputation (Rubin 1987) was shown to be effective to handle linkage ineligibles when analyzing a linked dataset between The National Health Interview Survey and CMS Medicare data (Zhang et al. 2016). This study used multiple imputation (MI) to handle missing data due to linkage ineligibility and missing claims for beneficiaries in Medicare Advantage programs to examine mammography screening. However, their study was not aimed at addressing inconsistencies in measuring program participation (e.g. “Medicaid undercount”) and did not compare the MI approach to the weight adjustment approach. An advantage to using MI is that all survey participants may be included in the analysis, minimizing the risk of bias. Moreover, the use of MI can account for nonresponse across other variables in the data.

The objective of this article is to compare three methods for determining Medicaid/CHIP status in health analyses of the NHANES-CMS Medicaid linked data: one that uses MI to impute the administrative Medicaid/CHIP status of those who are ineligible for linkage, a second that uses the linked data restricted to linkage eligible participants with a basic weight adjustment, and a third that uses self-reported Medicaid/CHIP status from the survey data. To compare these approaches, we examine associations between Medicaid enrollment and serum cotinine levels among children ages 3–15 years.

2 Materials 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 (Curtin et al. 2012). 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 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 survey question from 1999 to 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. This approach follows recommendations made by Klerman et al. (2012) when addressing similar concerns within analyses of the CPS.

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 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 the Children’s Health Insurance Program (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. It 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. However, as Klerman et al. (2012) concluded, it is a workable solution that helps to mitigate any Medicaid-CHIP confusion since the two are inseparable in the survey.

2.3 Data linkage

Data linkage between NHANES and the CMS MAX files is performed regularly by the National Center for Health Statistics’ Data Linkage program. Survey participants are linkage-eligible if they supply sufficient personally identifiable information, such as social security number (SSN) and health insurance claim number, and if their SSN is verified by the Social Security Administration’s Enumeration Verification System (Golden et al. 2015). Survey participants are ineligible for linkage if personally identifiable information is not provided. Records missing SSN on the Medicaid side are excluded as well. Linkage eligible survey participants whose SSN, month and year of birth, and sex exactly match the CMS MAX files were considered “linked”. The linkage between NHANES data and the CMS MAX files is complete for NHANES 1999–2004. Linked enrollment and claims data for NHANES 2005 through 2012 is expected to be available by the end of 2018. The currently linked data corresponds to all Medicaid/CHIP claims files between 1999 and 2009.

2.4 Motivating example: serum cotinine levels

To illustrate differences between methods, we examined serum cotinine levels. Serum cotinine is an important biomarker of children's health and has been shown to have disparate levels by race, age, income, and other demographic factors (EPA 2016; US Department of Health and Human Services 2006). Serum cotinine levels are used to estimate exposure to environmental tobacco smoke (ETS). Serum cotinine is a metabolite of nicotine and is preferred over nicotine as a surrogate of ETS due to its longer-half life (Benowitz 1996). Among nonsmokers, cotinine levels at or above the detection limit indicate secondhand smoke exposure within the previous 1–2 days (Schober et al. 2008).

According to the U.S. Surgeon General there is no safe level of exposure to ETS. Children and infants who are exposed to ETS have an increased risk of sudden infant death syndrome (SIDS), acute lower respiratory infection, middle ear disease, bronchitis, pneumonia, and asthma. There is also sufficient evidence to infer a causal relationship between exposure to ETS after birth and lower levels of lung function during childhood (Federal Interagency Forum on Child and Family Statistics 2016; US Department of Health and Human Services 2006). As a result, reducing the proportion of children exposed to second hand smoke is a Leading Health Indicator monitored for Healthy People 2020 (Office of Disease Prevention and Health Promotion 2014). In addition, ETS is monitored in America's Children and the Environment, the U.S. Environmental Protection Agency's report on data related to children's environmental health (Environmental Protection Agency 2016).

Recent research suggests that exposure to ETS through the form of third hand smoke (THS) disproportionately effects populations that spend time in multiunit housing and spaces with frequent occupancy change (Chambers et al. 2015, DeCarlo et al. 2018). THS consists of residual tobacco smoke pollutants that remain on surfaces and in dust after tobacco has been smoked. These pollutants are re-emitted into the gas phase or react with oxidants and other compounds in the environment to yield secondary pollutants. THS exposure results from the involuntary inhalation, ingestion, or dermal uptake of THS pollutants (Matt et al. 2011a, b). THS persists in smoker's homes and automobiles for months after occupancy, even after the vacant home or vehicle has been cleaned (Matt et al. 2008, 2011a, b). Moreover, central heating, ventilating, and air conditioning (HVAC) can move air contaminated by ETS throughout a building (Spengler 1999). Thus large apartment buildings with many occupants who smoke can increase the serum cotinine levels of children in that building, as can rented homes in which previous tenants smoked heavily (Matt et al. 2008, 2011a, b; King et al. 2013; Winickoff et al. 2010). As low family income is a criteria for Medicaid enrollment (Medicaid.gov) and low income families are more likely to live in multi-family and rental units (King et al. 2013), given the health consequences from ETS, examining ETS for children enrolled in Medicaid/CHIP may help program planning and services. Healthy People 2020 has reported that based on the survey response to health insurance status, children aged 3–11 years with public health insurance had more than twice the rate of secondhand exposure as compared to children with private health insurance (Office of Disease Prevention and Health Promotion 2014).

During physical examinations, blood was obtained by venipuncture for survey participants who were ≥ 1 year and serum cotinine was measured for those who were ≥ 3 year. Details about the measurement have been described elsewhere (Centers for Disease Control and Prevention 1999–2000). From 1999 to 2001 the laboratory limit of detection (LOD) for serum cotinine was 0.05 ng/mL and the below the LOD threshold was 0.035 ng/mL. From 2002 to 2004 the LOD was 0.15 ng/mL and the below the LOD threshold was 0.011 ng/mL. Values below the LOD were substituted with a constant of the detection limit divided by $\sqrt{2}$. Due to a highly skewed distribution, serum cotinine levels were logarithmically transformed for all imputations and analyses.

2.5 Analytic sample

Using NHANES 1999–2004, this study included children ages 3–15 years. Figure 1 indicates how many NHANES 1999–2004 participants ages 3–15 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 were fully or partially enrolled in Medicaid in the same state that their interview was conducted for at least 1 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 = 2632), those who were linkage eligible and not linked were classified as non-Medicaid/CHIP beneficiaries (n = 4190), and those who were ineligible for linkage had unknown Medicaid/CHIP status (n = 2034). To examine associations between Medicaid and serum cotinine levels, the sample was further restricted to nonsmoking children who participated in the MEC examination (n = 8514). Smokers were identified as children with serum cotinine levels > 10 ng/mL (Rebagliato 2002).

Medicaid/CHIP participation is also available as a survey response. As described below, however, survey response was used as a variable in the imputation models to inform the imputation for children ineligible for linkage and as the third strategy for determining Medicaid/CHIP enrollment status in the comparative analysis examining the association between Medicaid/CHIP enrollment and cotinine. It was not used to adjust Medicaid enrollment status defined by linkage to administrative records for linkage eligible children.

2.6 Medicaid assessment and approach for item non-response

Three alternative strategies were applied for determining Medicaid/CHIP enrollment status: the first used multiply imputed data, the second used weight adjusted NHANES CMS MAX linked data without imputation, and the third used survey data without imputation. Each strategy also used a different approach for handling missing information for cotinine and other covariates used in the estimation of the Medicaid/CHIP-cotinine association.

For the MI method, Medicaid/CHIP enrollment for linkage eligible children was defined using the classification above, as full or partial enrollment in the same state that their interview was conducted for at least 1 day during the month and year of their interview, and MI was used to impute enrollment status for children who were ineligible for linkage. MI was simultaneously used to impute missing information for cotinine and other covariates among all children (n = 8514).

Author Manuscript
Author Manuscript
Author Manuscript

For the weight adjustment method, Medicaid/CHIP enrollment for linkage eligible children was also defined as full or partial enrollment in the same state that their interview was conducted for at least 1 day during the month and year of their interview. However, children who were ineligible for linkage were dropped from the analysis for this method (n = 1912); among linkage eligible children an additional 1550 were dropped due to missing cotinine or other covariates. The basic weight adjustment post-stratified on race and Hispanic origin, age, and gender using the WTADJUST procedure in SAS.

Survey response was used to classify Medicaid/CHIP enrollment for the method using survey data. Children missing survey response information on Medicaid/CHIP status, cotinine values, or other covariate values were dropped from the survey response analysis (2268 cases dropped).

2.7 Multiple imputation

MI was conducted using SAS version 9.3 PROC MI (Fully Conditional Specification option) with 100 imputations. Survey variables that were thought to be related to either Medicaid/CHIP enrollment or second hand smoke exposure were considered as covariates. Design variables representing the primary sampling units (counties), strata, and sample weights were also incorporated. 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. This approach was used previously when imputing income for the National Health Interview Survey (Schenker et al. 2006).

In addition to the survey and design variables, a final explanatory variable was created by combining self-reported family income with Kaiser’s 2004 reports of Medicaid and S-CHIP state income thresholds (Henry J Kaiser Family Foundation). This variable, which classifies children as Medicaid eligible, S-CHIP eligible, or neither, was used exclusively as a predictor in the imputation model. 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 people might be eligible for Medicaid/CHIP.

Data were assumed to be missing at random (MAR) because there were a variety of covariates related to linkage eligibility that plausibly could explain the missing data mechanism. A series of nonresponse analyses regressing the missingness indicator, linkage eligibility, on the variables in the dataset indicated that the variables gender, age, race/ethnicity, country of birth, citizenship status, and survey Medicaid status, are all significant predictors of linkage ineligibility. This result is consistent with previously published literature (Bohensky et al. 2010; Carter et al. 2010). Furthermore, an inclusive imputation approach was used, including many covariates related to linkage eligibility and serum cotinine, so that the MAR assumption would become more plausible (Little and Rubin 2002).

The final set of predictors was selected so as to yield the lowest estimates of the fraction of missing information (Rubin 1987). 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 in the main analysis to fully incorporate the relationship among these variables as well as with aforementioned predictors (Collins et al. 2001). Table 1 lists the variables included in the imputation model and their level of missingness. Administrative Medicaid/CHIP status was missing for 23% of children (i.e. 23% of children were ineligible for linkage) and serum cotinine was missing for 22.5% of children.

2.8 Weight adjustment procedures

Weight adjustments were created using a model-based, calibration approach. For the reasons highlighted in the Sect. 2.7, data were assumed to be MAR. Sample weights were adjusted by post-stratifying to external population control totals. The fully saturated model included gender, age category, race/ethnicity, and all possible interactions among variables. These variables were chosen due to their propensity to both linkage eligibility and the underlying survey design as well as the availability of external population control totals. Age and race/ethnicity were categorized as they were for the creation of the original survey examination weights (age: ≤ 5 years, 6–11 years, 12–15 years, 16–19 years, 20–39 years, and 40–59 years; race: Mexican–American, other Hispanic, non-Hispanic white, non-Hispanic black, all other races and ethnicities including multi-racial). The marginal weight adjustment across groups was examined for large differences and high variability. All adjustments were performed using the PROC WTADJUST command in SAS. Summary statistics, correlations, and scatterplots of adjusted and unadjusted weights were inspected to examine cell counts and qualitatively identify outliers. No extreme weights were observed.

2.9 Analysis

Log linear models were fit to examine the association between serum cotinine levels and Medicaid/CHIP enrollment, while controlling for other sociodemographic and smoking exposure characteristics. Since the most appropriate comparison group could not be identified, Medicaid/CHIP enrollment was treated as binary: enrolled or not enrolled. Other covariates included gender (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 household interview (3–5, 6–11, 12–15 years), education of the household reference person (≤ High school graduate/GED, some college/associates degree/ college graduate or higher), poverty-income ratio (ordinal: ≤ 1, 1.01–2, 2.01–3, 3.01–4, > 4), and whether or not someone in the home smokes (yes/no). 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. The poverty income ratio 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.

With the exception of gender, all of these variables have been previously shown to be associated with both Medicaid/CHIP enrollment (Dubay and Kenney 1996; Kincheloe et al. 2007; Simon et al. 2013) and ETS (King et al. 2010; Orton et al. 2014, 2016).

All analyses were performed with SAS-callable SUDAAN, version 9.3 PROC REGRESS, 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 $p < 0.05$. Differences across the three methods were evaluated by comparing the β coefficients, the precision, and the relative standard error (RSE) associated with each β coefficient (defined as the standard error of the coefficient divided by the coefficient itself).

2.10 Sensitivity analyses

Two sensitivity analyses were conducted. The first aimed to separate the effect of imputing our primary variable of interest, Medicaid/CHIP, from imputing other covariates used in the regression models for the ETS-Medicaid/CHIP analyses. The second evaluated the imputation model.

To separate the effect of imputing our primary variable of interest, Medicaid/CHIP, from imputing other covariates used in the regression models for the ETS-Medicaid/CHIP analyses, the first sensitivity analysis imputed Medicaid/CHIP and cotinine (along with the other covariates) separately instead of simultaneously. To do this the imputation was performed in two steps: first Medicaid/CHIP was imputed and then (in a separate imputation) cotinine and all missing covariates (besides Medicaid) were imputed. For this sensitivity analysis, the imputation model for the Medicaid/CHIP imputation was the same as the model for the original analysis with one exception: it did not include the cotinine variable. The imputation model for cotinine and the other covariates was the same as the original with two exceptions. First, the administrative Medicaid/CHIP variable was not included. Second, the continuous PSU variable represented the average cotinine levels among children within each PSU instead of the average Medicaid/CHIP enrollment within each PSU.

These separately imputed data were applied to the models for Medicaid and cotinine using the three approaches for determining Medicaid status (MI, weight adjust, and survey response). For the MI method, the two imputations were combined and analyzed as though they were one imputation. For the weight adjustment method and the survey response method, only the second imputation (cotinine and all missing covariates) was incorporated into the analyses. As with the primary analysis, the MI method and the weight adjustment method defined Medicaid/CHIP enrollment (before imputation) based on the administrative files. A study participant was identified as a Medicaid/CHIP beneficiary if they were enrolled during the same state, month, and year as the interview. The survey response method based Medicaid/CHIP enrollment on survey response.

A second sensitivity analysis assessed the imputation model. After limiting the data to those who were linkage eligible, the administrative Medicaid enrollment status for a random subset of 23% of children was set to missing and the imputation model used to reclassify

Medicaid enrollment. This 23% corresponded to the percentage of children in the original sample who were missing administrative Medicaid enrollment status because they were ineligible for linkage. The exercise was done over 100 imputations and the average number of concordant classifications across all imputations was compared to the average number of discordant classifications.

3 Results

3.1 Main analysis

Table 2 summarizes the characteristics of the sample by the variables used in the regression analysis and by both linkage eligibility and administrative Medicaid/CHIP status. From a total of 8856 children, 80% (n = 6822) were linkage eligible and 20% (n = 2034) were ineligible for linkage (all percentages weighted). Among those who were linkage eligible, 30% (n = 2632) were identified as Medicaid/CHIP beneficiaries and 70% (4190) were not. Among those who were both linkage eligible and enrolled in Medicaid/CHIP, 69% identified as Medicaid/CHIP beneficiaries within the survey, 36% were exposed to second hand smoke in the home, and the geometric mean for serum cotinine was 0.32 ng/mL. Among those who were linkage eligible and not enrolled in Medicaid/CHIP, 95% identified as non-Medicaid beneficiaries within the survey, 20% were exposed to second hand smoke in the home, and the geometric mean for serum cotinine was 0.09 ng/mL.

Among those who were eligible for linkage, some demographic differences were observed between children enrolled in Medicaid/CHIP and those who were not. Medicaid/CHIP beneficiaries displayed a higher percentage of children who identified as non-Hispanic black (28% vs. 9%), a higher percentage of children whose household reference person had less than or equal to a high school education (72% vs. 41%), a higher percentage of children who were identified as having a poverty income ratio ≤ 2 (91% vs. 33%), and a higher percentage of children who lived with someone who smoked inside the home (36% vs. 20%).

In addition, children who were ineligible for linkage more closely matched the linkage eligible children who were non-Medicaid beneficiaries than the linkage eligible children who were Medicaid beneficiaries in terms of socioeconomic status and smoking characteristics. The geometric mean for the poverty income ratio was 0.82 (SE = 0.03) among linkage eligible Medicaid beneficiaries, 2.41 (SE = 0.07) among linkage eligible non-Medicaid beneficiaries, and 1.95 (SE = 0.08) among children ineligible for linkage. The geometric mean for serum cotinine was 0.32 (SE = 0.03) among linkage eligible Medicaid beneficiaries, 0.09 (SE = 0.01) among linkage eligible non-Medicaid beneficiaries, and 0.11 (SE = 0.01) among children ineligible for linkage.

The results from the primary analysis examining Medicaid/CHIP status and log serum cotinine levels across the three different methods for determining Medicaid/CHIP status are presented in Table 3. As can be seen by comparing the unadjusted regression to the full model, adjusting for covariates substantially affected the Medicaid/CHIP coefficients across all three methods. Using the adjusted models, after exponentiation Medicaid/CHIP beneficiaries had an increase in average serum cotinine levels of 40.5% (CI: 19.7, 64.9) based on the MI method, 40.5% (CI: 20.9, 63.2) based on the weight adjusted method, and

15.0% (CI: − 0.01, 32.3) based on the survey response method. At the $\alpha = 0.05$ level, the increase in log-cotinine levels for Medicaid/CHIP beneficiaries as compared to non-Medicaid/CHIP beneficiaries was statistically significant for both the MI method and the weight adjusted method, but not for the survey response method ($p = 0.0001$, 0.0001, and 0.07, respectively). The relative standard error (standard error/estimate) associated with the Medicaid/CHIP enrollment coefficient was 24% based on the imputation method, 21% based on the weight adjustment method, and 50% based on the survey response method.

The results associated with other covariates were consistent with previous literature (Spanier et al. 2015; Kit et al. 2013; Kaufmann et al. 2010) and consistent across approaches. Based on the MI model, cotinine levels were higher among children ages 3–5 years compared to children ages 12–15 years ($p = 0.0001$), higher among non-Hispanic blacks compared to non-Hispanic whites ($p = 0.06$), lower among Mexican–Americans compared to non-Hispanic whites ($p < 0.0001$), higher among children whose household reference person had less than or equal to a high school education compared to children whose household reference person had more than a high school education ($p < 0.0001$), and higher among children exposed to secondhand smoke in the home compared to children with no smoke exposure in the home ($p < 0.0001$). Associations between cotinine levels and PIR values were also statistically significant; as compared to children with a PIR value greater than 4, cotinine levels were higher among all other PIR categories, with p values ranging from 0.0003 to < 0.0001. All of the statistically significant coefficients for covariates in the weight-adjusted model were within 20% of the coefficients in the MI method.

3.2 Sensitivity analyses

Sensitivity analyses generally supported the robustness of the primary results. When cotinine and the other covariates were imputed separately from Medicaid/CHIP, the magnitude of the association between Medicaid/CHIP and cotinine decreased slightly for both the MI method and the weight adjusted linked data method. However the two methods were similar to one another in terms of RSE. After exponentiation, Medicaid/CHIP beneficiaries were associated with an increase in average serum cotinine levels of 33.6% (CI: 13.9, 56.8) based on the MI method, 37.7% (CI: 15.0, 63.2) based on the weight adjustment method, and 18.5% (CI: 2.0, 37.7) based on the survey response method (Table 3). The associated RSE for each of these methods was 28, 28, and 41%, respectively.

Limiting the sample to children who were linkage eligible decreased the sample size to 6822. The Medicaid/CHIP status of 1553 children (23%) was set to missing for the MI validation analysis. Across all 100 imputations, children were accurately identified as Medicaid/CHIP versus non-Medicaid/CHIP beneficiaries 79.9% of the time (n ≈ 1250 per imputation, range = 1216–1278), Medicaid/CHIP beneficiaries were identified as non-Medicaid/CHIP beneficiaries 10.2% (n ≈ 160 per imputation, range = 133–186) of the time, and non-Medicaid/CHIP beneficiaries were identified as Medicaid/CHIP beneficiaries 9.9% (n ≈ 154 per imputation, range = 129–182) of the time.

4 Discussion

We explored three methods that can be used to determine Medicaid/CHIP enrollment status when examining health outcomes within the NHANES data: the MI method, the weight adjustment method, and the survey response method. The estimated percent increases in average serum cotinine levels among Medicaid/CHIP beneficiaries were very similar using the MI method and the weight adjusted method. In fact, the point estimates were identical at 40.5%. The estimated percent increase based on the survey response method was 15%, resulting in a relative difference of 62.9% between the survey response method and the other two methods. Under the assumption that the administrative information is more accurate than survey report, these results suggest that the survey response method produced biased results.

The beta coefficients for the weight adjustment method and the MI method were identical ($\beta = 0.34$) and the confidence interval from the weight adjustment method was similar to, if slightly narrower than, the confidence interval from the MI method [MI: (0.18, 0.50), WA: (0.19, 0.49)]. In theory, multiple imputation would be expected to be more precise than weight adjustment for addressing linkage eligibility since an analysis using the MI includes all survey participants and not just the linkage eligible. We did not find this result in our comparison.

This analysis suggests that when using the NHANES CMS-MAX linked data, both the MI approach and the weight adjustment approach can be appropriate and effective ways to address the biases that may result from some survey participants being ineligible for linkage. It is unclear whether one method is preferable over the other and identifying a preferred method is difficult. Both methods display advantages. The MI approach enables researchers to include survey participants who are ineligible for linkage in statistical analyses that use the NHANES CMS MAX linked data. The MI method is also able to incorporate a large number of covariates related to both linkage eligibility and the outcome of interest, increasing the plausibility of the MAR assumption. In contrast, while age, race/ethnicity, and sex were sufficient for the weight adjustment method in this example, other outcomes may require a larger variety of covariates in order to ensure the MAR assumption is plausible. A large number of covariates for the weight adjustment method could lead to difficulties with model convergence (Judson et al. 2013). However, in this example, the weight adjustment approach produced an estimate with a similar, if smaller RSE and similar, if narrower, confidence interval than either of the other two methods.

The linked data files enhance the utility of the survey data by adding otherwise unavailable information from the administrative CMS-MAX files to the NHANES survey data. However, unlike the survey data, which was created for health outcomes research, the administrative files were created to monitor the progress of health care delivery by tracking enrollment, services, and costs. Differences in these two objectives lead to some inconsistencies between the data sources that ultimately provide limitations to the methods described here. As previously discussed, the NHANES survey question did not allow for a distinction to be made between Medicaid and CHIP beneficiaries and not all S-CHIP data are included in the CMS-MAX files. While the survey response to Medicaid/CHIP would, in theory, capture all children enrolled in these programs in all states, irregardless of whether they are S-CHIP or M-CHIP enrollees, the administrative linked file misses S-CHIP children for some states. Therefore for the MI and weight adjustment methods, a linkage eligible child who was enrolled in a non-reported S-CHIP program will have an incorrectly inferred false positive. Some of the differences between the linked data analyses and the survey response analysis could be due to this S-CHIP concern, but this is unlikely given what we know about the Medicaid undercount. If many S-CHIP enrollees were included in the survey count, but not in the linked data count, this would in fact attenuate the effect of the Medicaid undercount. Moreover, an undercount of S-CHIP enrollees would have a larger effect on estimations of totals or research on program participation than on estimates of association or health outcomes among children in these programs, which is the focus of this research. In our analyses we aimed to mitigate the effects of this inconsistency by using the combined Medicaid/CHIP category as suggested by Klerman et al. (2012) and by controlling for PSU (which do not cross state lines) and state income threshold in the imputation model.

Another difference in construct between the two datasets stem from fluctuations into and out of Medicaid eligibility. Among low-income beneficiaries eligibility fluctuations occur often (program churning) due to changes in employment or due to procedural reasons: families losing coverage during renewal periods, inadequate coordination between Medicaid and CHIP agencies, or changes in disability status or mental illness diagnosis (Orzol et al. 2015). While this is important information for tracking Medicaid enrollment, health outcomes researchers are more interested in distinguishing among those who have been traditionally enrolled in Medicaid versus those who have not. Some children who were enrolled in Medicaid for at least 1 day during the month of the survey may not be regularly or continuously enrolled, while other children who are typically enrolled in Medicaid may not have been during the exact month of the survey. This could lead to either the Medicaid or the non-Medicaid population being misinterpreted for research. It is important to note, however, that this is a limitation that stems from combining two very different datasets into one and using one specific time point, such as time of interview, in order to classify a child’s Medicaid status; it is not a limitation of either the imputation approach or the weight adjustment approach. Other limitations associated with the linked data files have been detailed elsewhere (Golden et al. 2015).

Finally, the generalizability of the results is unknown. We do not know whether the MI method and the weight adjustment method would perform similarly if used to examine other health outcomes. In this illustration, both serum cotinine and Medicaid/CHIP enrollment values were missing 23% of the time. If several major variables were missing values with random missingness patterns, then the MI method may have been more precise since it is used to simultaneously impute all missing variables, while the weighting can only be applied to handle missing data on one variable and thus leaves cases with other missing variables excluded.

Likewise, we do not know whether our conclusions about these methods would change if they were extended to later, more recent data (2005 onward). NHANES changed the structure of the health insurance questionnaire in 2005. The new structure distinguishes between Medicaid/M-CHIP beneficiaries and S-CHIP beneficiaries within the survey data, enabling comparisons with the linked data to be drawn across Medicaid/M-CHIP beneficiaries exclusively. This would address some of the misclassification concerns described above. Furthermore, the provision of SSN (necessary for data linkage) varies by survey year and by respondent characteristics (Golden et al. 2015). If the percentage and/or characteristics of children who are ineligible for linkage is markedly different between the survey years 1999–2004 and the survey years 2005–2012, then the imputation model may perform differently when used on the more recent data.

5 Conclusions

This analysis demonstrates that both multiple imputation and weight adjustment procedures are accessible and effective ways to incorporate survey participants who are ineligible for linkage into statistical analyses of the NHANES CMS MAX linked data files. It is unclear whether one method is preferable over the other, as this example produced similar results across methods. The advantage of both the MI and the weight adjustment methods is that they avoid using the possibly misreported Medicaid status from the survey data (Davern 2007; Davern et al. 2009). The advantage of the MI method over the weight adjustment method is that it incorporates all survey participants into the analysis and is able to include information from a large number of covariates. Based on this example, the weight adjustment method produced similar, if slightly more precise, estimates, but this may not be the case for other health outcomes which require adjusting for a larger number of covariates in order to produce unbiased estimates.

Acknowledgments

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

Benowitz NL: Cotinine as a biomarker of environmental tobacco smoke exposure. John Hopkins Univ. Sch. Hyg. Public Health Epidemiol. Rev 18(2), 188–203 (1996)

Bohensky MA, Jodlley D, Sundararajan V, Evans S, Pilcher DV, Scott I, Brand CA: Data linkage: a powerful research tool with potential problems. BMC Health Serv. Res 10(1), 346 (2010). 10.1186/1472-6963-10-346 [PubMed: 21176171]

Call KT, Davidson G, Davern M, Nyman R: Medicaid undercount and bias to estimates of uninsurance: new estimates and existing evidence. Health Serv. Res 43(3), 901–914 (2008). 10.1111/j.1475-6773.2007.00808.x [PubMed: 18546545]

Carter K, Shaw C, Hayward M, Blakely T: Understanding the determinants of consent for linkage of administrative health data with a longitudinal survey. Kòtuitui N. Z. J. Soc. Sci. Online 5(2), 53–60 (2010)

Centers for Disease Control and Prevention: Laboratory Procedure Manual, Cotinine, Serum, ID HPLC-APCI MS/MS. https://wwwn.cdc.gov/Nchs/Data/Nhanes/2003-2004/LabMethods/l06_c_met_cotinine.pdf (2004). Accessed 1 Sept 2017

Chambers C, Sung H, Max W: Home exposure to secondhand smoke among people living in multiunit housing and single family housing: a study of California adults, 2003–2012. J. Urban Health Bull. N. Y. Acad. Med 92(2), 279–290 (2015). 10.1007/s11524-014-9919-y

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

Curtin LR, Mohadjer LK, Dohrmann SM, 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. Vital Health Stat. 2 (155) (2012)

Davern M: 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. Census Bureau. https://www.census.gov/did/www/snacc/docs/SNACC_Phase_I_Full_Report.pdf (2007). Accessed 1 Sept 2017

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 Serv. Res 44(23), 965–987 (2009). 10.1111/j.1475-6773.2008.00941.x [PubMed: 19187185]

DeCarlo PF, Avery AM, Waring MS: Thirdhand smoke uptake to aerosol particles in the indoor environment. Sci. Adv (2018). 10.1126/sciadv.aap8368

Dubay LC, Kenney GM: The effects of Medicaid expansions on insurance coverage of children. Future Child. 6(1), 152–161 (1996). 10.2307/1602499 [PubMed: 8689258]

Environmental Protection Agency: America's Children and the Environment, Biomonitoring: Cotinine. https://www.epa.gov/ace/ace-biomonitoring-cotinine (2016). Accessed 1 Sept 2017

Federal Interagency Forum on Child and Family Statistics: America's Children in Brief: Key National Indicators of Well Being, 2016. https://www.childstats.gov/pdf/ac2016/ac_16.pdf (2016). Accessed 1 Sept 2017

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 and Medicaid Services. Vital Health Stat. 1 (58), 8–10, 14–20 (2015)

Henry J Kaiser Family Foundation: State Health Facts. http://kff.org/state-category/medicaid-chip/ Accessed 17 May 2018

Judson DH, Parker JD, Larsen MD: Adjusting sample weights for linkage-eligibility using SUDAAN. National Center for Health Statistics http://www.cdc.gov/nchs/data/datalinkage/adjusting_sample_weights_for_linkage_eligibility_using_sudaan.pdf (2013). Accessed 1 Sept 2017

Kaufmann RB, Babb S, O'Halloran A, Asman K, Bishop E, Tynan M, Caraballo RS, Pechacek TF, Bernert JT, Blount B: Vital signs: nonsmokers' exposure to secondhand smoke—United States, 1999–2008. Morb. Mortal. Wkly. Rep 59(35), 1141–1146 (2010)

Kincheloe J, Frates J, Brown ER: Determinants of children's participation in California's Medicaid and SCHIP programs. Health Serv. Res 42(2), 847–866 (2007) [PubMed: 17362221]

King BA, Travers MJ, Cummings KM, Mahoney MC, Hyland AJ: Secondhand smoker transfer in multiunit housing. Nicotine Tob. Res 12(11), 1133–1141 (2010). 10.1093/ntr/ntq162 [PubMed: 20889473]

King BA, Babb SD, Tynan MA, Gerzoff RB: National and state estimates of secondhand smoke infiltration among US multiunit housing residents. Nicotine Tob. Res 15(7), 1316–1321 (2013). 10.1093/ntr/nts254 [PubMed: 23248030]

Kit BK, Simon AE, Brody DJ, Akinbami LJ: US prevalence and trends in tobacco smoke exposure among children and adolescents with asthma. Pediatrics 131(3), 407–414 (2013) [PubMed: 23400612]

Klerman J, Plotzke MR, Davern M: CHIP reporting in the CPS. Medicare Medicaid Res. Rev (2012). 10.5600/mmrr.002.03.b01

Little RJA, Rubin DB: Statistical Analysis with Missing Data, 2nd edn. Wiley, New York (2002)

Matt GE, Quintana PJ, Hovell MF, Chatfield D, Ma DS, Romero R, Uribe A: Residual tobacco smoke pollution in used cars for sale: air, dust, and surfaces. Nicotine Tob. Res 10(9), 1467–1475 (2008). 10.1080/14622200802279898 [PubMed: 19023838]

Matt GE, Quintana PJ, Destaillats H, Gundel LA, Sleiman M, Singer BC, Peyton J, Benowitz N, Winickoff JP, Rehan V, Talbot P, Schick S, Samet J, Wang Y, Hang B, Martins-Green M, Pankow JF, Hovell MF: Thirdhand tobacco smoke: emerging evidence and arguments for a multidisciplinary research agenda. Environ. Health Perspect. 119(9), 1226–2018 (2011a)

Matt GE, Quintana PJE, Zakarian JM, Fortmann AL, Chatfield DA, Hoh E, Uribe AM, Melbourne FH, et al.: When smokers move out and nonsmokers move in: residential third hand smoke pollution and exposure. Tob. Control 20, e1 (2011b). 10.1136/tc.2010.037382

Medicaid.gov: Keeping America Healthy. https://www.medicaid.gov/medicaid/eligibility/index.html Accessed 17 May 2018

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. Natl. Health Stat. Rep 72, 1–9 (2014)

Office of Disease Prevention and Health Promotion: Healthy People 2020 Leading Health Indicators: Environmental Quality. https://www.healthypeople.gov/sites/default/files/HP2020_LHI_Environ_Qual.pdf (2014). Accessed 17 May 2018

Orazine CI, Arias WA, Magee SR, King E: Non-smoking pregnant woman and their fetuses are exposed to environmental tobacco smoke as a result of living in multiunit housing. J. Expo. Sci. Environ. Epidemiol (2016). 10.1038/jes.2016.34

Orton S, Jones LL, Cooper S, Lewis S, Coleman T: Predictors of children's secondhand smoke exposure at home: a systematic review and narrative synthesis of the evidence. PLoS ONE 9(11), e112690 (2014). 10.1371/journal.pone.0112690 (Accessed 1 September 2017) [PubMed: 25397875]

Orzol SM, Hula L, Harrington M: Program churning and transfers between Medicaid and CHIP. Acad. Pediat 15(3s), S56–S63 (2015). 10.1016/j.acap.2015.02.006 (Accessed 1 September 2017)

Rebagliato M: Validation of self reported smoking. J. Epidemiol. Community Health 56, 163–164 (2002) [PubMed: 11854332]

Rubin DB: Multiple Imputation for Nonresponse in Surveys. Wiley, New York (1987)

Schenker N, Raghunathan TE, Chiu P, Makuc DM, Zhang G, Cohen AJ: Multiple imputation of missing income data in the national health interview survey. J. Am. Stat. Assoc 101(475), 924–933 (2006). 10.1198/016214505000001375

Schober SE, Zhang C, Brody DJ, Marano C: Disparities in secondhand smoke exposure—United States, 1988—1994 and 1999–2004. Morb. Mortal. Wkly. Rep 57(27), 744–747 (2008)

Simon AE, Driscoll A, Gorina Y, Parker JD, Schoendorf KC: A longitudinal view of child enrollment in Medicaid. Pediatrics132, 656–662 (2013). 10.1542/peds.2013-1544 [PubMed: 24062367]

Spanier AJ, Beck AF, Huang B, McGrady ME, Drotar DD, Peake RWA, Kellogg MD, Kahn RS: Family hardships and serum cotinine in children with asthma. Pediatrics 135(2), e416–e423 (2015). 10.1542/peds.2014-1748 [PubMed: 25583915]

Spengler JD: Buildings operations and ETS exposure. Environ. Health Perspect 107(Suppl2), 313–317 (1999) [PubMed: 10375293]

The Kaiser Commission on Medicaid and the Uninsured: Medicaid A Primer: Key Information on the Nation's Health Coverage Program for Low-Income People. Report No. 7332–05. https://kaiserfamilyfoundation.files.wordpress.com/2010/06/7334-05.pdf (2013). Accessed 1 Sept 2017

Truffer CJ, Wolfe CJ, Rennie KE: 2015 Actuarial Report on the Financial Outlook for Medicaid. Office of the Actuary, Centers for Medicare and Medicaid Services, United States Department of Health and Human Services https://www.medicaid.gov/medicaid/financing-and-reimbursement/downloads/medicaid-actuarial-report-2015.pdf (2015). Accessed 1 Sept 2017

US Department of Health and Human Services, CDC: The Health Consequences of Involuntary Exposure to Tobacco Smoke: A Report of the Surgeon General—Executive Summary. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Coordinating Center for Health Promotion, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health (2006)

Winickoff JP, Gottlieb JD, Mello MM: Regulation of smoking in public housing. N. Engl. J. Med 362, 2319–2325 (2010). 10.1056/NEJMhle1000941 [PubMed: 20554988]

Zhang G, Parker J, Schenker N: Multiple imputation for missingness due to nonlinkage and program characteristics: a case study of the National Health Interview Survey linked to medicare claims. J. Surv. Stat. Methodol 4(3), 319–338 (2016). 10.1093/jssam/smw002

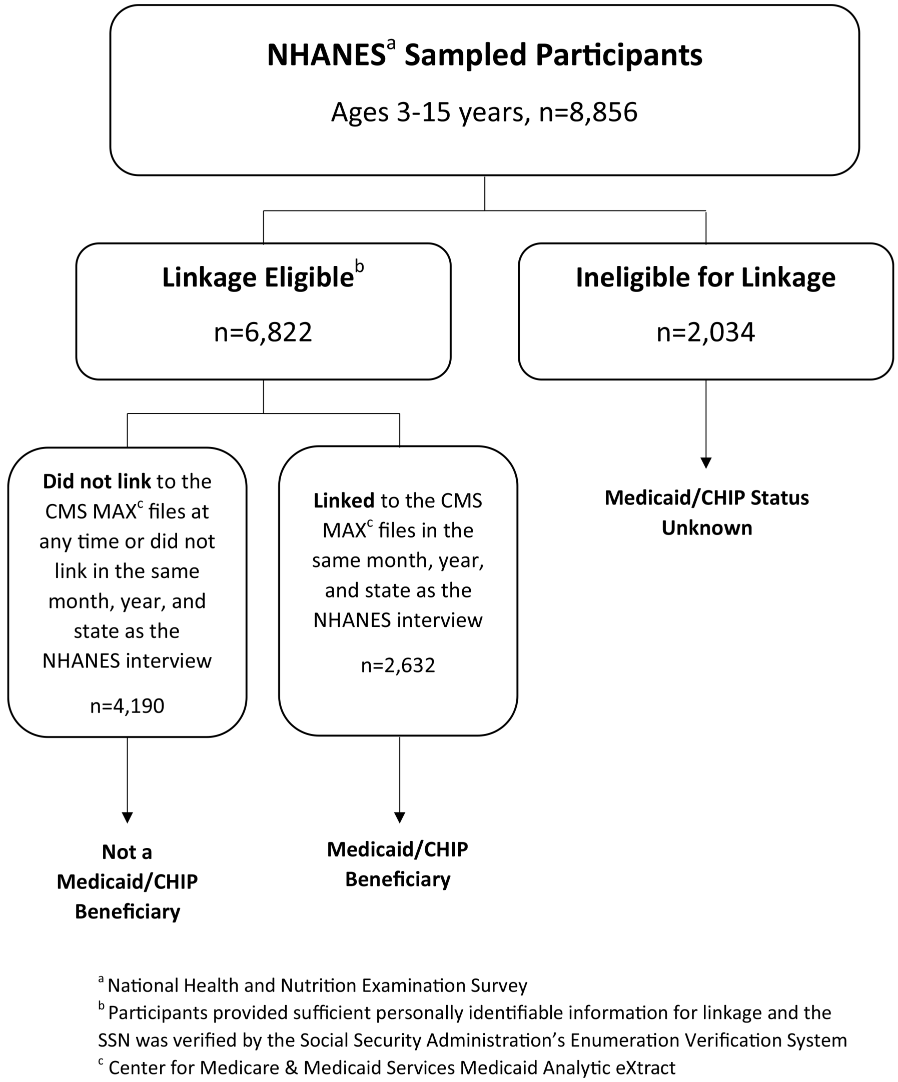

Fig. 1.

Medicaid classification based on administrative data. CDC/NCHS National Health and Nutrition Examination Survey data (1999–2004) linked to Centers for Medicare and Medicaid Services' Medicaid eXtract files (1999–2007)

Table 1

Variables included in the final imputation model and their level of missingness

	n	% missing	Imputation method
Demographic variables			
Gender	0	0	Logistic regression
Age (in years)	0	0	Linear regression
Race/ethnicity	0	0	Discriminant function
Nativity	1	0.01	Discriminant function
Citizenship status	21	0.24	Discriminant function
Poverty income ratio	761	8.59	Logistic regression
Age of household reference person (in years)	6	0.07	Linear regression
Nativity of household reference person	314	3.55	Discriminant function
Education level of household reference person	350	3.95	Logistic regression
Subject specific variables			
Self-reported general health condition	3	0.03	Logistic regression
Average number of health care visits last year	11	0.12	Logistic regression
Home ownership status	146	1.65	Discriminant function
Indicator: mother smoked while pregnant	118	1.33	Logistic regression
Average # of cigarettes/day in the home[a]	233	0.03	Linear regression
NHANES survey medicaid status	160	1.81	Discriminant function
Qualify for medicaid, SCHIP, or neither based on self-reported income and state eligibility thresholds	761	8.59	Discriminant function
Variables of interest			
Log serum cotinine (ng/mL)	1988	22.45	Linear regression
Administrative medicaid status	2034	22.97	Logistic regression
Design based variables			
Primary sampling units	0	0	Linear regression
Strata	0	0	Discriminant function
Mobile examination center sample weights	0	0	Linear regression
Total	8856		

National Health and Nutrition Examination Survey data, 1999–2004

Linked with Centers for Medicare and Medicaid Services Medicaid Analytic eXtract files, 1999–2007

Sampled participants aged 3–15 years

All percentages are unweighted

[a]Children who did not live with a smoker were coded as 0 cigarettes/day

Table 2

Characteristics of the study sample by linkage eligibility and administrative Medicaid status

Characteristic	Eligible for linkage			Ineligible for linkage
	Medicaid	Non-Medicaid	Medicaid and non-Medicaid Combined	Medicaid status Unknown
n	2632	4190	6822	2034
Gender[b]				
Male	0.54	0.52	0.52	0.46
Female	0.46	0.48	0.48	0.54
Age category[ab]				
3–5 years	0.26	0.22	0.23	0.19
6–11 years	0.49	0.47	0.48	0.43
12–15 years	0.25	0.31	0.29	0.38
Race/ethnicity[a]				
Mexican American	0.13	0.11	0.12	0.15
Non-hispanic white	0.44	0.68	0.61	0.55
Non-hispanic black	0.28	0.09	0.15	0.16
All other ethnicities including multiracial	0.15	0.12	0.13	0.15
Education of household reference person[a]				
Less than high school/high school graduate/GED	0.72	0.41	0.51	0.48
Some college/AA/college graduate or higher	0.28	0.59	0.49	0.53
Poverty income ratio[a]				
≤ 100%	0.60	0.10	0.25	0.21
100% < PIR ≤ 200%	0.31	0.23	0.25	0.24
200% < PIR ≤ 300%	0.06	0.22	0.17	0.18
300% < PIR ≤ 400%	0.02	0.16	0.12	0.15
400% < PIR ≤ 500%	0.02	0.30	0.22	0.23
Someone in the home smokes[ab]				
Yes	0.36	0.20	0.25	0.19
No	0.64	0.80	0.75	0.81
Survey medicaid response[ab]				
Enrolled	0.69	0.05	0.24	0.13
Not enrolled	0.31	0.95	0.76	0.88
Serum cotinine (ng/mL)[ab]	0.98 (0.08)	0.47 (0.05)	0.62 (0.06)	0.48 (0.07)
Serum cotinine (ng/mL) (geometric)	0.32 (0.03)	0.09 (0.01)	0.14 (0.01)	0.11 (0.01)

National Health and Nutrition Examination Survey (1999–2004) linked to Centers for Medicare and Medicaid Services Medicaid Analytic eXtract files (1999–2009)

Sampled participants aged 3–15 years who participated in the MEC examination

Medicaid status based on NHANES CMS MAX linked data: matched on state, year, and month of interview

Categorical variables: weighted proportions

Continuous variables: weighted means (standard error)

All variables described are included in the adjusted regression models

Survey medicaid response is included in the adjusted model for the survey response method only *1* Springer

PIR is an index for the ratio of self-reported family income divided by the federal poverty guideline specific to family size, year, and state

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

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

Table 3

Multiple linear regression analyses of log serum cotinine levels among MEC examined children aged 3–15 years

Medicaid enrollment status	Multiple imputation method			Weight adjustment method			Survey response method		
	β (SE)	*p* value	Sample size	β (SE)	*p* value	Sample size	β (SE)	*p* value	Sample size
Unadjusted regression									
Medicaid beneficiary	1.18 (0.10)	<0.0001	8514	1.22 (0.10)	<0.0001	5424[a]	1.08 (0.12)	<0.0001	5312[b]
Not a medicaid beneficiary	0.00 (0.00)	-		0.00 (0.00)	-		0.00 (0.00)	-	
Adjusted regression									
Medicaid beneficiary	0.34 (0.08)	0.0001	8514	0.34 (0.07)	0.0001	5052[a]	0.14 (0.07)	0.07	6246[b]
Not a medicaid beneficiary	0.00 (0.00)	-		0.00 (0.00)	-		0.00 (0.00)	-	
Sensitivity analysis: Medicaid and cotinine imputed separately									
Medicaid beneficiary	0.29 (0.08)	0.001	8514	0.32 (0.09)	0.0007	6602[a]	0.17 (0.07)	0.03	8396[b]
Not a medicaid beneficiary	0.00 (0.00)	-		0.00 (0.00)	-		0.00 (0.00)	-	

National Health and Nutrition Examination Survey (1999–2004) linked to Centers for Medicare and Medicaid Services Medicaid Analytic eXtract files (1999–2009)

All analyses were restricted to nonsmokers who had a physical examination

All beta coefficients correspond to a binary Medicaid enrollment variable: Medicaid beneficiaries versus Non-Medicaid beneficiaries

[a]Medicaid status defined by administrative data. Exclusion criteria: Ineligible for linkage, nonresponse to cotinine or covariate values

[b]Medicaid status defined by survey response. Exclusion criteria: nonresponse to survey reported Medicaid status, cotinine, or other covariates