

# Enhancing the Medical Expenditure Panel Survey Data on Medicaid Beneficiaries through Data Linkage

Lisa B. Mirel<sup>1</sup> and Jeffrey M. Gonzalez<sup>1</sup>

<sup>1</sup>Division of Statistics, Research and Methods, Center for Financing, Access and Cost Trends, Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850

## Abstract

The MEPS Household Component (MEPS-HC) is a complex, multi-stage, nationally representative sample of the U.S. civilian non-institutionalized population. Each year a new sample is drawn as a subsample of households from the prior year's National Health Interview Survey (NHIS) conducted by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC). The MEPS-HC and the NHIS integrated design enable the indirect linkage of some MEPS-HC participants to Medicaid administrative records which offers innovative opportunities to assess survey quality and improve imputation methodology.

This paper looks at outcomes based on MEPS participants that linked to administrative records, using data from the NCHS data linkage program. Focusing on the Medicaid Analytic eXtract (MAX) files, we analyzed dental visits. We describe the utility of linked data for validating survey reporting, assessing expenditure estimates, and improving imputation methods in MEPS-HC. We also discuss plans for future work in this area.

**Key words:** Medical Expenditure Panel Survey (MEPS), National Health Interview Survey (NHIS), Medicaid, Data Linkage, Dental

## 1. Background<sup>1</sup>

The MEPS-HC is a complex, multi-stage, nationally representative sample of the U. S. civilian non-institutionalized population. It is an annual survey of about 14,000 households, conducted since 1996. There are five rounds of data collection that cover a two year reporting period. It provides national estimates on health care use, expenditures, insurance coverage, sources of payment, access to care and health care quality.

Each year a new MEPS-HC sample is drawn as a subsample of households that participated in the prior year's NHIS conducted by the NCHS, CDC. The NHIS is a multi-purpose health survey that serves as the principal source of information on the health status and health behaviors of the civilian, non-institutional U.S. population. NHIS uses a complex, multi-stage sample design, oversampling Asians, Hispanics and blacks (Division of Health Interview Statistics, 2011). This complex survey design carries over to the MEPS-HC through the set of NHIS responding households that comprise the frame for MEPS-HC sample selection.

---

<sup>1</sup> The views expressed in this paper are those of the authors and no official endorsement by the Department of Health and Human Services or the Agency for Healthcare Research and Quality is intended or should be inferred.

The MEPS-HC and the NHIS integrated design enable the indirect linkage of some MEPS-HC participants to their Medicaid administrative records through the NCHS data linkage program. The NCHS program links various surveys (including NHIS) with mortality data from the National Death Index (NDI), enrollment and claims records from the Centers for Medicare & Medicaid Services (CMS), and benefit history data from the Social Security Administration (SSA).

For the MEPS-HC, the indirect linkage to the CMS Medicaid claims data offers an opportunity to validate respondent reports of Medicaid coverage and utilization of health care services received and to explore possible ways to improve the imputation of missing payment information for Medicaid recipients. Data collected in the MEPS-HC on expenditures for Medicaid beneficiaries are extremely limited because financial transactions typically occur directly between the provider/plan and the state Medicaid agency. In other words, the survey respondent typically doesn't have any information on how much was paid by Medicaid. Although health care events in MEPS are strictly based on reports of health care utilization in the HC, linkage to data collected in the MEPS Medical Provider Component (MPC) is the primary vehicle for obtaining Medicaid expenditures for physician visits and hospital care. Other types of services (i.e. event types) however, are not included in the MPC due to budgetary constraints. These event type categories are:

- Office-based care not provided or directly overseen by an MD (e.g. physical therapists),
- Dental care,
- Non-Agency home health care (i.e. paid independent providers) and
- Other medical equipment and supplies.

Because data are not collected in the MPC for these services, it's necessary to impute all Medicaid expenditure data for these non-MPC event types. In general, the imputation approach used is to adjust a total charge value (which may be reported in the HC or imputed) by a payment-to-charge ratio which varies by event type. This paper focuses on dental care for Medicaid recipients. We will assess the accuracy of reported Medicaid coverage and utilization and assess the imputation of Medicaid expenditures for those with dental events.

## **2. Methods**

To assess the accuracy between household reported Medicaid coverage and household-reported number of dental visits we relied on the linkage of MEPS participants and the CMS MAX files. We also used the linked data to assess the imputation procedure of Medicaid expenditures for dental visits by comparing means and medians of the imputed Medicaid payments in the MEPS-HC and the MAX claims data. The MAX data are a set of person-level files on Medicaid eligibility, service utilization, and payments. The dataset files are created to support research and policy analysis. According to CMS, there are limitations that should be considered when analyzing the MAX data. The MAX data only contain services and payment data for beneficiaries during periods of Medicaid eligibility, and these data are limited to beneficiaries not enrolled in managed care and children without CHIP coverage. (Simon, et al., 2014). In addition, there is substantial variation in Medicaid at the state level, including program eligibility, services offered, and provider reimbursement (Bouchery, Brief 9, 2012) and (Simon, et al., 2014).

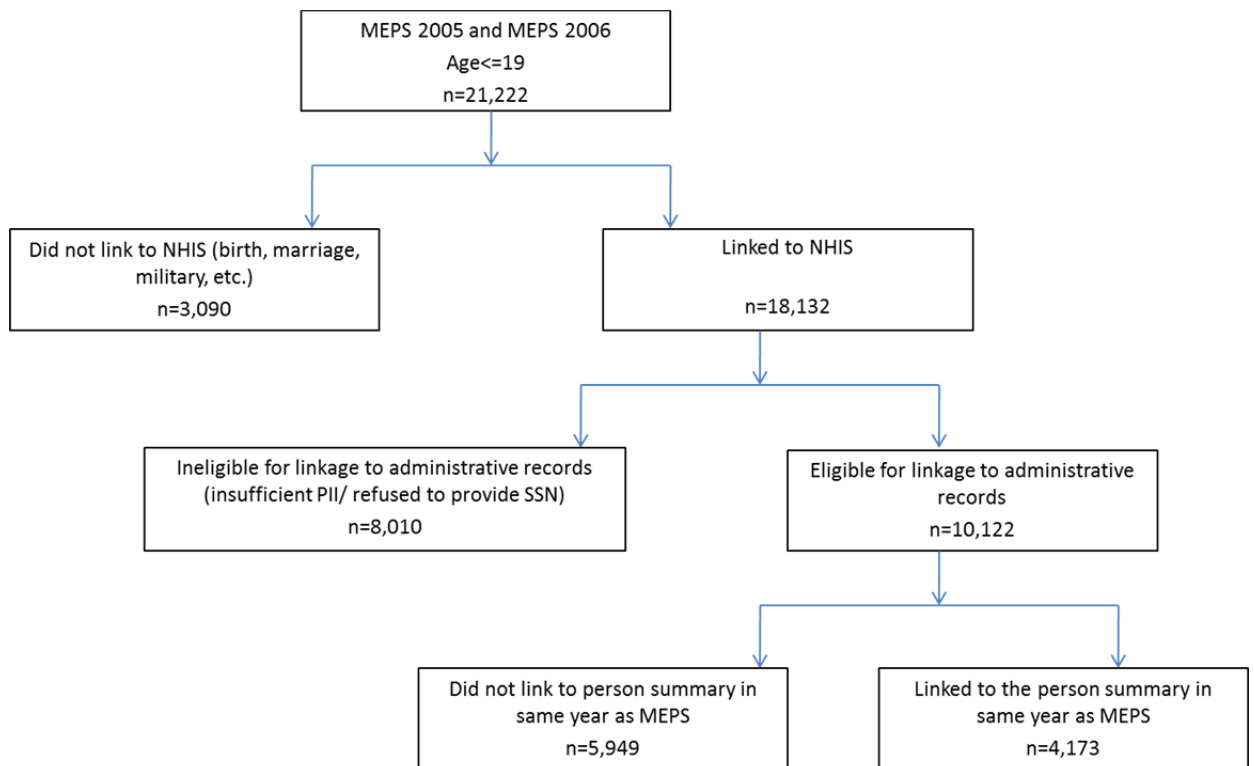
This paper focuses on children 19 years of age and younger. We limited the analysis to children since children enrolled in Medicaid are entitled to preventive dental services (Bouchery, Brief 9, 2012). In MEPS 2005 and MEPS 2006, there were 21,222 children 19 years of age and younger. Of these, 8,496 (46%) had a dental visit. Since dentists are not included in the MPC, the payments for the dental visits were not captured. As a result, the expenditures for the dental visits were imputed for our sample of survey participants with Medicaid coverage. Of the 8,496 children with dental visits, 3,439 children (24%) had Medicaid coverage and therefore had imputed Medicaid expenditures for the dental visits.

## **2.1 Data Linkage**

The data linkage was part of an interagency agreement between NCHS, CMS, SSA, and the U.S. Department of Health and Human Services Assistant Secretary for Planning and Evaluation (ASPE). The linked data are available in the NCHS Research Data Center for linkage eligible respondents. NHIS participants were defined as linkage eligible if the survey respondent had provided sufficient personally identifiable information (PII) and/or had not refused to provide their Social Security Number (SSN) during the NHIS interview. The SSN was then verified by the SSA Enumeration Verification System. Linkage eligible respondents were linked with their Medicaid data using the following identifiers: SSN, date of birth (month and year) and gender. A successful link matched on all three identifiers (Simon, et al., 2014). Then CMS extracted data for the linked MEPS-HC participants from the Medicaid files, including the Person Summary (PS) and the Other Services (OT) files (Simon, et al., 2014). The 1999-2009 MAX files were linked to the NHIS 1994-2005 for linkage eligible respondents. In order to use the most recent years of data, we analyzed MEPS 2005 and MEPS 2006 data (which corresponds to NHIS 2004 and 2005).

## **2.2 Sample**

Of the 21,222 participants less than 20 years of age in MEPS 2005 and MEPS 2006, 10,122 linked to NHIS and were deemed linkage eligible based on the information obtained during the NHIS interview (Figure 1). Not all MEPS participants linked to NHIS and therefore were not part of the analytic sample. Newly-eligible people join the MEPS sample households after the NHIS is conducted in the previous year, usually due to several reasons, including marriage, birth and returning home from the military (Chowdhury, Wun, & Machlin, 2012). Of the 10,122 linkage eligible participants, 4,173 linked to MAX PS file in the same year as MEPS.

**Figure 1.** Sample breakdown

### 2.3 Accuracy in reporting of Medicaid

First we assessed accuracy of reporting Medicaid as a source of insurance coverage by comparing the number of people identified as having Medicaid/CHIP in MEPS (as reported by the household respondent) to those that linked to the MAX Person Summary (PS) file, using all linkage eligible participants (n=10,122). The PS file includes person level data on eligibility, demographics, basis of eligibility, managed care enrollment, and a summary of utilization and Medicaid payment by type of service. The file contains one record for every individual enrolled for at least one day during the year.

To define Medicaid coverage in MEPS we used the household questionnaire responses. In the MEPS-HC the respondent was asked:

“Has anyone in the family been covered by {Medicaid/{STATE NAME FOR MEDICAID}} or {STATE CHIP NAME} at any time {since (START DATE)/between (START DATE) and (END DATE)}?”

If the respondent answered “Yes” to the above question, they were then asked:

“Who is covered {Medicaid/{STATE NAME FOR MEDICAID}} or {STATE CHIP NAME}?”

If Medicaid/CHIP was reported for a participant in at least one of the rounds then that MEPS participant was identified as having Medicaid/CHIP coverage for at least some part of the survey year. The participants identified as having Medicaid/CHIP coverage were then compared to the PS file. Concordance was defined as 1) concordant no: those that did not report Medicaid/CHIP in MEPS and were not found on the MAX PS and 2)

concordant yes: those that reported Medicaid/CHIP as an insurance source in at least one round of data collection and were on the MAX PS file in the same year as the survey.

## 2.4 Accuracy of reporting of dental visits and payments

To assess the accuracy of reporting dental visits and to assess the imputation for Medicaid expenditures for dental visits, we focused on the 4,173 MEPS-HC participants that linked to the MAX PS file. We identified dental visits in MEPS using the Full Year Consolidated file for 2005 and 2006. This file includes the total number of dental visits for the year and the total dental payments by Medicaid for the year. In order for a dental visit to be counted in the analysis there had to be a total dental payment by Medicaid greater than zero for the year.

The number of dental visits and payment information for Medicaid beneficiaries were derived using the MAX Other Services (OT) file. The MAX OT file contains claim records for all non-institutional Medicaid services, including dental services, physician services, lab/x-ray, clinic services and premium payments. Each claim record includes diagnosis, procedure code, date of service, payment and charge. The OT file includes a record for each claim. In order to make the data more comparable to how events are reported in the MEPS-HC, we created one record per service begin date. We then summed across the claims for each beneficiary to get the total Medicaid payment for the calendar year.

We calculated an overall concordance for dental visits from the two files. Concordance was defined as 1) concordant no: not reporting a dental visit in MEPS-HC or having a dental visit but no payment by Medicaid for a dental visit in the year in the MEPS-HC and no dental claims on the OT file (n=2,277) and 2) concordant yes: those who reported at least one dental visit in MEPS-HC and had a dental payment by Medicaid greater than zero for the year and had at least one dental claim on the MAX OT file (n=821).

## 2.5 Analysis

Dichotomous variables indicating Medicaid coverage and whether or not the MEPS-HC participant had a dental visit were computed, using both data sources. Agreement for Medicaid coverage and dental visits was assessed between the two sources of data, MEPS-HC and MAX, using the Kappa statistic. The Kappa statistic corrects the percentage of agreement between raters (or data source in our analysis) by taking into account the proportion of agreement expected by chance (Cohen, 1960).

We also assessed the accuracy of the mean number of household reported dental visits and the mean imputed Medicaid payments per year in MEPS-HC compared to MAX. NCHS recommends adjusting survey weights for linkage eligibility (Judson, Parker, & Larsen, May 2013). This recommendation is applicable for Medicare beneficiaries over 65 since Medicare coverage is guaranteed for individuals aged 65 and older who have worked and paid into the system. However, this recommendation may not be applicable for Medicaid beneficiaries. Unlike the Medicare population, linkage eligibility in our analysis does not imply Medicaid eligibility since Medicaid is a means-tested program and not all children under age 20 are Medicaid eligible. Appropriate weights have not been developed yet for this subset of linked MEPS participants and MAX data. Future work will be done to develop weights for this subset.

We calculated the median, inter-quartile range (IQR) and mean (standard error (SE)) for the number of dental visits and Medicaid payments per year from both sources of data.

Paired t-tests were used to compare mean visit counts and payments. All significance tests were evaluated using a two-sided  $p$ -value  $<0.05$  as the level of statistical significance. Since all of the Medicaid payments were imputed for dental visits, the comparison of mean payments per year was used to assess the effectiveness of the MEPS imputation procedure. All statistical analyses were performed using SAS (version 9.3 SAS Institute Inc., Cary, N.C.).

### 3. Results

#### 3.1 Accuracy in reporting of Medicaid coverage

The overall concordance for the accuracy in reporting Medicaid as a source of insurance coverage was 88% (Table 1). The Kappa statistic was 0.77 which represents substantial agreement (Landis & Koch, 1977). This concordance rate is similar to results from a multi-phase research project referred to as the Medicaid Undercount project that was conducted to help explain why discrepancies exist between survey estimates of enrollment in Medicaid and the number of enrollees reported in state and national administrative data (SHADAC, ASPE, AHRQ, NCHS, CMS, & U.S. Census Bureau, 2010).

For those who did not report Medicaid/CHIP in MEPS-HC and were on the MAX PS file ( $n=312$ ), we looked to see what other types of insurance coverage they reported in the MEPS-HC interview. Some of the discordance may come from misreporting the type of insurance coverage. Of the 312 reporting no Medicaid coverage in MEPS-HC but were found in the PS file, 171 (55%) reported private health insurance in at least one round of MEPS, 46 (15%) reported some other public health insurance and 95 (30%) did not report any health insurance coverage in any of the rounds of the MEPS interview.

A possible explanation for the other discordant group (the 843 MEPS-HC participants identified as having Medicaid/CHIP in MEPS but were not found on the PS) is that service information in MAX may be missing or incomplete for those with managed care plans and for children with CHIP coverage (Simon, et al., 2014).

**Table 1.** Accuracy in reporting of Medicaid for children less than 20 years of age, MEPS 2005 and MEPS 2006 linked to 2005 and 2006 MAX files

Medicaid/CHIP in MEPS-HC	On the MAX PS file in the same year as MEPS-HC		Total
	No	Yes	
No	50.4% $n=5,106$	3.1% $n=312$	53.5 $n=5,418$
Yes	8.3% $n=843$	38.1% $n=3,861$	45.4 $n=4,704$
Total	58.8% $n=5,949$	41.2% $n=4,173$	100% $n=10,122$

#### 3.2 Accuracy in reporting of dental events and imputed Medicaid payments

First we compared the dichotomous variable, indicating at least one dental visit in the survey year from MEPS-HC and from MAX, for the 4,173 MEPS-HC participants that were on the person summary file. The overall concordance was 74.2% (Table 2). The Kappa statistic was 0.41 which represents moderate agreement (Landis & Koch, 1977).

The discordance between those who did not report a dental visit in MEPS but had a claim in the MAX OT (n=461) could be due to under-reporting of health care events in the household survey (Zuvekas & Olin, 2009). The discordance for those who have a dental visit reported in MEPS-HC but did not have a claim in MAX OT file (n=614) may be due to differential reporting by states or incomplete MAX data for those in a managed care plan or CHIP.

**Table 2.** Accuracy in reporting of dental visits for children less than 20 years of age, MEPS 2005 and MEPS 2006 linked to 2005 and 2006 MAX files

Have one or more dental visit with a Medicaid payment in MEPS-HC	Have one or more dental claim in MAX OT file		Total
	No	Yes	
No	54.6% n=2,277	11.1% n=461	65.7% n=2,738
Yes	14.7% n=614	19.6% n=821	34.3% n=1,435
Total	69.3% n=2,891	30.7% n=1,282	100% n=4,173

Next we compared the average number of visits in MEP-HC and MAX and average payment per year for the 4,173 participants (Table 3). There were no statistically significant differences between the mean number of dental visits reported in MEPS-HC and the mean number of visits in MAX. The difference for the mean payments per year between the two sources was marginally significant ( $p < 0.06$ ).

**Table 3.** Number of dental visits and Medicaid payments for children less than 20 years of age, MEPS 2005 and MEPS 2006 linked to 2005 and 2006 MAX files

n=4,173	MAX	MEPS
Visits		
Median (IQR)	0 (0-1)	0 (0-1)
Mean (SE)	0.65 (0.02)	0.65 (0.02)
Payment per year		
Median (IQR)	0 (0-\$71)	0 (0-\$52)
Mean (SE)	\$94.7 (4.17)	\$80.6 (6.91)

We then compared the mean number of visits and payments for those that were in the concordant yes group (had one or more dental visit in MEPS with an imputed yearly Medicaid payment greater than zero and had one or more dental claim in MAX). This group of concordant yes was about 20% of the total sample that linked to the PS file (n=821). As noted above, all of the Medicaid payments in MEPS were imputed for this group. The difference in the mean number of visits between the two sources was statistically significant ( $p\text{-value} < 0.0001$ ) among the concordant yes group (MAX number of visits was higher). Similarly, the difference in the payments per year was statistically significant ( $p\text{-value} < 0.05$ ), on average the Medicaid dollar amounts from the administrative records was higher than what was imputed in MEPS (Table 4).

**Table 4.** Number of dental visits and Medicaid payments for the concordant yes group in the linked sample of children less than 20 years of age, MEPS 2005 and MEPS 2006 linked to 2005 and 2006 MAX files

n=821	MAX	MEPS
Visits		
Median (IQR)	2 (1-3)	1 (1-2)
Mean (SE)	2.3 (0.06)	1.9 (0.05)
Payment per year		
Median (IQR)	\$185 (94-385)	\$87 (50-174)
Mean (SE)	\$326.7 (15.41)	\$258.0 (31.39)

The last set of analyses looked at the summary statistics for the discordant group (Table 5). This includes 1) those that had a claim in MAX but did not have any dental visits reported in MEPS-HC or did not have a yearly dental Medicaid payment greater than zero (n=461) and 2) those that did not have a claim in MAX but did have a dental visits reported with a dental Medicaid payment greater than zero (n=614). Even though these two groups are not directly comparable we note that the number of visits looks similar between the two sources but the payments are greater in the MAX file.

**Table 5.** Number of dental visits and Medicaid payments per year for the discordant group in the linked sample of children less than 20 years of age, MEPS 2005 and MEPS 2006 linked to 2005 and 2006 MAX files

	MAX n=461	MEPS n=614
Visits		
Median (IQR)	1 (1-2)	1 (1-2)
Mean	1.8 (0.06)	1.8 (0.07)
Payment per year		
Median (IQR)	\$158 (76-296)	\$79.5 (52-152)
Mean (SE)	\$275.1 (16.47)	\$203.1 (17.65)

#### 4. Summary

In summary, the overall mean number of visits was similar from the two data sources (0.65) and the mean Medicaid payments per year were slightly higher in the MAX file (this difference was marginally significant). For those in the concordant yes group the difference in the mean number of visits was statistically different from the MAX number of visits which was about 20% higher than the MEPS-HC reported number of visits. This may be due to under-reporting in the household survey (Zuvekas & Olin, 2009). The mean payment per year in MAX was also about 20% higher than the MEPS-HC payment per year. This could be a result of how the Medicaid data are imputed for the MEPS-HC. We are currently working on ways to improve the imputation, using the linked data to assess charges and the payment to charge ratios which are used in the imputation process for non-MPC Medicaid events.

In addition, our future work will include exploring factors associated with differences in payment for the concordant yes group. We are planning to run similar analyses using more up to date data when the next linkage occurs at NCHS. In addition, we are exploring comparisons to other surveys.



## Acknowledgements

The authors would like to thank Sadeq Chowdhury, Joel Cohen, Steve Cohen, Steve Machlin, D.E.B. Potter, Marc Roemer, Nina Verevkina, and Sam Zuvekas for their insightful and helpful contributions to this paper. In addition, the analyses for this project could not have been completed without the help of Frances McCarty and Adrienne Jones of the NCHS RDC.

## References

- Bouchery, E. (Brief 9, 2012). *Utilization of Dental Services Among Medicaid Enrolled Children*. Princeton, NJ: Mathematica Policy Research.
- Chowdhury, S., Wun, L-M., & Machlin, S. (2012). Linking Medical Expenditure Panel Survey to the National Health Interview Survey: Weighting and Estimation Considerations. *Joint Statistical Meetings, Section of Survey Research Methods* (pp. 4435-4449). Alexandria, VA: American Statistical Association.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.
- Division of Health Interview Statistics. (2011, June 15). *2010 NHIS Survey Description Document*. Retrieved December 13, 2013, from NCHS CDC: [ftp://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/Dataset\\_Documentation/NHIS/2010/srvydesc.pdf](ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2010/srvydesc.pdf)
- Ezzati-Rice, T. M., Rohde, F., & Greenblatt, J. (2008). *Sample Design of the Medical Expenditure Panel Survey Household Component, 1998–2007*. Agency for Healthcare Research and Quality. Rockville, MD: Methodology Report No. 22.
- Judson, D. H., Parker, J. D., & Larsen, M. D. (May 2013). *Adjusting sample weights for linkage-eligibility using SUDAAN*. Hyattsville, Maryland: National Center for Health Statistics.
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics, 33*(1), 159-174.
- SHADAC, ASPE, AHRQ, NCHS, CMS, & U.S. Census Bureau. (2010). *Report for Phase VI of the SNACC Medicaid Undercount project: Comparing MSIS and MEPS*. The University of Minnesota : The State Health Access Data Assistance Center (SHADAC).
- Simon, A. E., Driscoll, A. K., Golden, C., Tandon, R., Duran, C. R., Miller, E. A., et al. (2014). *Documentation and Analytic Guidelines for NCHS surveys linked to Medicaid Analytic eXtract (MAX) files*. Hyattsville, MD: National Center for Health Statistics.

Zuvekas, S. H., & Olin, G. L. (2009). Validating Household Reports of Health Care Use in the Medical Expenditure Panel Survey. *Health Services Research, 44*(5 Part I), 1679-1700.