Linking Medical Expenditure Panel Survey to the National Health Interview Survey: Weighting and Estimation Considerations[1]

Sadeq Chowdhury, Lap-Ming Wun, and Steve Machlin
Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850

Abstract

The Medical Expenditure Panel Survey (MEPS) is an annual survey with an overlapping panel design where the sample for a panel is drawn from the responding households of the previous year's National Health Interview Survey (NHIS). Because of this relationship between the two surveys, data from the NHIS can be linked to the MEPS to expand the survey's analytic capacity. However, as the MEPS is conducted a year after conducting the NHIS, not all persons in a MEPS sample can be linked with the NHIS sample due to the joining of new persons in some households. Options for analyzing such a linked dataset are to exclude the cases with missing NHIS data and apply the original MEPS weight, exclude the cases with missing NHIS data and apply an adjusted MEPS weight, or impute missing values so that the full dataset can be analyzed using the original MEPS weight. This paper presents the results of an investigation on the importance of weighting adjustments for analysis and estimation when MEPS is linked with the NHIS full, sample adult or sample child files. For the analysis, the MEPS 2007 FY file is linked with the NHIS 2005 and 2006 files and the MEPS weights are adjusted for non-linkage. The estimates produced with adjusted and unadjusted weights are then compared to assess the impact of an adjustment of MEPS weight for non-linkage. The analysis shows that when MEPS is linked with the NHIS full file the impact of a weighting adjustment is negligible, but when linking with the NHIS sample adult or sample child file the impacts on many estimates are significant and a weighting adjustment for non-linkage is strongly recommended.

Key Words: Data Linkage, Linking Surveys, MEPS, NHIS, Weighting

1. Introduction

The Medical Expenditure Panel Survey (MEPS), conducted by the Agency for Healthcare Research and Quality (AHRQ), provides nationally representative estimates of health care use, expenditures, sources of payment, and health insurance coverage for the U.S. civilian non-institutionalized population. The Household Component of MEPS is the core survey which is based on a complex national area probability sample design. A new sample panel is selected every year and is followed for two consecutive years; two overlapping panels are combined each year to produce annual estimates from a total sample of about 14,000 households and 30,000 individuals. The National Health Interview Survey (NHIS) is a large multi-purpose health survey conducted by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), and is the principal source of information on the health status and health behaviors of the civilian, non-institutionalized, household population of the United States. The NHIS sample size varies in the range of about 29,000 to 39,000 households

[1] The views expressed in this paper are those of the authors and no official endorsement by the Department of Health and Human Services (DHHS) or the Agency for Healthcare Research and Quality (AHRQ) are intended or should be inferred.

(about 75,000 to 100,000 persons) from year to year. The MEPS is directly related to the NHIS because MEPS samples are selected from responding households to the previous year's NHIS. Because of the relationship between the two surveys, data from the NHIS can be linked to the MEPS to expand MEPS's analytic potential. Most persons responding to the MEPS can be linked back to the NHIS for longitudinal analysis with the same or similar variables or for cross-sectional analysis with additional variables from the NHIS. Cohen (2010) discusses the integration of MEPS and NHIS data for some expanded analytic possibilities.

The MEPS person-level file can be linked with the NHIS core file, the NHIS sample adult file and the NHIS sample child file. The NHIS sample adult file is based on a random sub-sampling of one adult per family and the NHIS sample child file is based on a random sub-sampling of one child per family (where one or more children under 18 years are present) from the NHIS core sample. The NHIS core module collects health, demographic, and other related information about everyone in the family while the sample adult and sample child modules collect more specific data on health status, health care services, and behavior of adults and children respectively. By linking the MEPS records with the NHIS records in these files, the analytic capacity of MEPS or NHIS data can be enhanced in several ways, such as follows:

a) Cross-sectional analyses of MEPS data can be enriched by appending some additional background-type NHIS variables which are not collected in the MEPS, such as home ownership, citizenship status, and nativity status, or some health conditions. For example, Bernard et al. (2011) conducted a study using the cancer status of the linked MEPS respondents from the NHIS.

b) Retrospective analysis of MEPS data can be done by checking the status of the cases (say with a certain condition) a year or two earlier at the time of conducting NHIS. For example, cases in the MEPS sample can be linked to NHIS and analyzed according to whether or not they had cancer at the time of NHIS.

c) Prospective analysis of the NHIS cases can be done by checking the subsequent status of the linked cases a year or two later at the time of conducting the MEPS.

d) Longitudinal analysis of the linked cases can be done by using the variables that are common in both NHIS and MEPS. Since a MEPS panel is interviewed five times to gather data for a two-year period, the longitudinal analysis of MEPS data is constrained to two years, but by linking and adding more information from the prior year's NHIS, the duration of a longitudinal analysis can be extended for some variables. Rhoades and Cohen (2012) conducted such a study to analyze the long-term uninsured population.

One of the issues that needs to be addressed in analyzing a MEPS file linked with NHIS is that not all persons in the MEPS final sample with positive weights can be linked with the NHIS sample, even though the MEPS sample was selected from NHIS responding households. This is because newly-eligible[2] persons join MEPS sample households after conducting the NHIS due to marriage, migration, or birth. Moreover, since the NHIS sample adult and sample child files only include a subsample of the full NHIS sample, the non-linkage rates of the MEPS cases to NHIS sample adult and sample child files are even higher. The non-linkage rate is 5-6% for the core NHIS file but around 55% for the sample adult and the sample child files. Excluding these non-linked cases without making a proper adjustment to the MEPS sampling weight may introduce a

[2] Persons who are eligible at the time of NHIS and join MEPS households subsequently are considered to be *non-key* in MEPS and receive a zero weight.

bias to the estimates obtained from the linked file. Similar to non-response bias, the extent of non-linkage bias will depend on the relationship between non-linkage and the analytic variables of interest. Ideally, in analyzing such a linked dataset, the MEPS or NHIS weights of the linked cases should be appropriately adjusted to account for the non-linked cases. If the non-linkage is random without any association with target variables, the bias due to non-linkage would be ignorable and a simple overall weight adjustment would be appropriate. Also when the non-linkage rate is very low, the relevant missing observations of the non-linked cases due to non-linkage could simply be treated as missing, as is the case for many other survey variables for which there are missing values due to non-imputed item non-response. One advantage of being able to perform the analysis of a MEPS-NHIS linked file using the usual MEPS weight is that it avoids the need for AHRQ or data users to develop a special weight variable for linked analyses.

In this study, we assess the extent of non-linkage bias in the absence of a weighting adjustment. More specifically, we investigate if non-linkage is associated with any socio-economic determinants of target variables and discuss the issues associated with deriving appropriate weights to produce estimates from a linked file. Then we assess the impact of weight adjustment by comparing estimates produced using non-adjusted weights with estimates produced using weights adjusted for non-linkage for a range of MEPS and NHIS variables. The analysis was designed to help inform a decision about the need for routinely producing an adjusted analytic weight for linked analysis.

Since the cases that should be included in a linked analysis and whether the MEPS or NHIS weight (adjusted or unadjusted) is most appropriate depend on the analytic objective(s), in this investigation we adjust weights and compare the estimates for analysis types a) and b) described above i.e. linked analyses that enhance the utility of MEPS (using MEPS weights) by adding variables from the NHIS full, sample adult, or sample child files.

2. Linking MEPS 2007 Full-Year (FY) File with NHIS

A MEPS full year (FY) file consists of two panels – an old panel (year 2) and a new panel (year 1). For example, the MEPS 2007 FY file, which we will use for this analysis, consists of Panels 11 and 12. As shown in Table 1 below, Panel 11 was selected in 2006 from the 2005 NHIS and Panel 12 was selected in 2007 from the 2006 NHIS. Therefore, if a MEPS FY file is linked with the NHIS, the persons from the old panel will be linked with a 2-year earlier NHIS file and the new panel will be linked with a 1-year earlier NHIS file.

Table 1: Relationship between MEPS panels in 2007 FY file and 2005 and 2006 NHIS annual samples

2005	2006	**2007**
		MEPS 2007 FY File
NHIS 2005 ==>	MEPS Panel 11: Yr 1=>	MEPS Panel 11: Yr 2
	NHIS 2006 ==>	MEPS Panel 12: Yr 1

Table 2 shows the distribution of persons by link status and the corresponding link rates for the 2007 MEPS FY file when linked with the NHIS core (full) file, sample adult and sample child files. The overall unweighted link rates of MEPS are 95.8% with the NHIS full file, 43.9% with the sample adult file and 42.9% with the sample child file,

while the overall weighted link rates are 95.7% with the core file, 45.8% with the sample adult file and 45.9% with the sample child file.

Table 2. Distributions of persons in 2007 MEPS FY file by panel and link status with 2005-06 NHIS files

MEPS Panel		NHIS								
		Full File (2+ Years)[a]			Sample Adult (18+ Years)[a]			Sample Child (2-19 Years)[a,b]		
		Not Linked	Linked	Total	Not Linked	Linked	Total	Not Linked	Linked	Total
11	Number	752	15,151	15,903	6,550	5,067	11,617	2,785	2,021	4,806
	%	4.7	95.3	100.0	56.4	43.6	100.0	57.9	42.1	100.0
12	Number	449	12,187	12,636	5,156	4,100	9,256	2,133	1,669	3,802
	%	3.6	96.4	100.0	55.7	44.3	100.0	56.1	43.9	100.0
Total	Number	1,201	27,338	28,539	11,706	9,167	20,873	4,918	3,690	8,608
	%	4.2	95.8	100.0	56.1	43.9	100.0	57.1	42.9	100.0
Total	Weighted %	4.3	95.7	100.0	54.2	45.8	100.0	54.1	45.9	100.0

[a]All ages as of the MEPS poststratification reference date of 12/31/07. [b]Age group 2-19 years is used since most of 18-19 year olds in MEPS were 16-17 years at the time of NHIS

Table 2 and all subsequent analyses exclude children less than 2 years old (at the time of 2007 MEPS i.e., on 12/31/2007), because most were born after the 2005 and 2006 NHIS was conducted and a link is not expected. In other words, we considered these cases out-of-scope for a linked analysis. Similar to linkage rates of persons less than 2 years, the link rate with the sample adult file is also low for 18-19 years olds because most of these individuals were under 18 years at the time of NHIS. However, 18-19 years were not excluded from the analysis as these cases did exist at the time of NHIS and may be reasonably represented by 20-29 years cases. Also, individuals aged 18-19 years are potentially linkable to the sample child file because most of these cases were less than 18 years at the time of NHIS and thus could be part of the sample child file.

To give an idea of the likely impact of excluding non-linked cases from any analysis without proper weighting adjustment, Table 3 presents weighted distributions of the linked cases by various background characteristics typically used in the nonresponse and poststratification/raking adjustments of the MEPS sample. When linking with the NHIS full file, the link rates are high and do not show much variation. For the older age groups (>65 years) and for the White/other groups the link rate was slightly higher, and for the uninsured and the poor groups the link rate was slightly lower. When linking with the sample adult and sample child files, the variation in link rates appear to be slightly higher. The link rates for both of these files decreases rapidly with the increase in the number of adults/children in the family, as would be expected due to the selection of one adult/child from each family for sample adult/child modules. For the sample adults, the link rate is the highest for the elderly since they are more likely to live in single adult households and lowest for the 18-29 age groups, because of the shifting of 18-19 years' from child to adult status as discussed above. Linkage rates were also high for black and white/other race/ethnicity groups. On the other hand, the link rate is slightly lower for the middle/high income adults and tends to vary by health insurance status. For the sample child link, the rates appear to vary to some extent by poverty status and health insurance

status. In summary, other than number of adults/children, the link rate does not seem to vary dramatically by most of the characteristics considered.

Table 3. Link rates by selected characteristics for linking the 2007 MEPS FY file with 2005-06 NHIS files

Characteristic	Category	Weighted Link Rate		
		Full File	Sample Adult	Sample Child
Overall		95.7	45.8	45.9
Age category	2-4	83.5	-	47.2
	5-9	95.3	-	44.9
	10-17	95.4	-	45.5
	18-29	93.3	33.8	47.4
	30-44	96.7	47.0	-
	45-64	97.7	47.1	-
	65-74	98.1	54.1	-
	75+	97.8	60.3	-
Sex	Male	95.4	42.7	46.4
	Female	96.6	48.8	45.3
Race/Ethnicity	Hispanic	94.6	37.6	42.9
	Black	93.6	47.9	43.0
	Asian	92.7	36.0	46.2
	White/Others	96.5	47.7	47.5
Poverty Status	Poor	92.7	50.5	37.4
	Near Poor/Low Inc	95.7	50.2	42.3
	Middle/High Income	96.2	44.1	49.4
MSA Status	MSA	96.5	47.2	45.3
	Non-MSA	95.6	45.6	46.0
Health Insurance	Private	96.7	45.8	48.5
	Public	94.0	51.8	40.8
	Uninsured	93.5	39.8	44.3
Panel	11	95.3	45.7	45.2
	12	96.2	46.0	46.6
Number of Persons/Adults/ Children in the Family	1	97.5	75.5	76.0
	2	96.5	41.6	42.4
	3	94.6	26.5	29.5
	4	95.0	18.8	22.3
	5+	94.5	14.8	17.5

3. Weighting of Linked Sample

Whether the MEPS or NHIS weights should be adjusted to derive a linked weight depends on the objective of the analysis. If the base of the analysis is the MEPS reference year as in analysis types a) and b) discussed in the introduction, then the adjusted weight should represent the cases that are in-scope in the MEPS reference year. On the other hand if the base of the analysis is the NHIS as in analysis type c), then the adjusted weight of the linked cases should represent the population in the NHIS reference period. Again if the objective of the analysis is to perform a longitudinal analysis as in analysis type d) then the weight should represent all in-scope persons during the whole period of the longitudinal analysis.

As mentioned above, we conduct the investigation concentrating on the analysis type a) and b) and adjust the MEPS weight of the linked cases. The in-scope persons at the end of the MEPS year are adjusted to the same CPS-based population control total used for the MEPS 2007 FY weighting.

Since the linked dataset has non-linked cases that by definition have missing values for the NHIS variables, an ideal analysis of the linked dataset will either require imputation of the missing NHIS variables or an adjustment to the MEPS weight to account for the missing cases. The advantage of imputation is that the full file can be used without excluding the non-linked cases that have non-missing data for many MEPS variables, but the disadvantage is that sound imputation approaches may be difficult to develop. Moreover, imputation may be more appropriate when linking with the NHIS full file where the non-link rate is only 4-5%, but may not be a reasonable option for linking with the sample adult and sample child files because over half of MEPS cases cannot be linked to these files.

We adjust the MEPS weight for linked analysis such that the weight represents the population in the MEPS reference year. Different weights are derived for linking with the NHIS full file, the sample adult file and the sample child file. Similar to the non-response adjustment, adjustments to the MEPS weights of the linked cases are made to account for the non-linked cases to derive the weights for a linked analysis. We refer to the adjusted weight as 'link weight' and the estimates based on this weight as 'adjusted estimates'.

Similar to the MEPS weighting procedures, we made separate adjustments to the weights of the cases in-scope and out-of-scope at the end of the MEPS year. Since the number of out-of-scope cases on December 31 is very small (less than 1%), a simple poststratification adjustment of weights is made for these cases. The poststratification cells are formed using reasons for out-of-scope (Decedent, Nursing Home Entrant, and Other) by panel. For the sample adult and the sample child files, the adjustments to the weights for out-of-scope cases are made by forming similar poststratification cells but with some collapsing of cells due to small sample sizes.

For the in-scope cases, link weights are derived by applying raking adjustments to accommodate most variables used in the raking adjustment to derive MEPS FY weights (see Machlin et al. 2010). For the full file, the poststratification/raking adjustment is applied directly to the MEPS weights. However, for the sample adult and child link files an initial adjustment is made to the MEPS weight for number of adults/children in the family before applying the raking adjustment because only one adult/child per family was selected in NHIS. Consequently, the selection probability varied considerably and the adults or children in families with the smaller number of adults/children are over represented in the linked sample adult/child files (see Table 1). This initial adjustment is made at the overall level to ensure that the sums of the initial weights by number of adults/children (1, 2, 3+) in the link sample adult or sample child file is the same as the corresponding sums of the weights in the MEPS full file. This adjustment increases the weights of the linked persons in 2 and 3+ adult/child family so that persons from these families are also appropriately represented. The persons with 3 or more adults/children in the family are combined to avoid too much variation in the resulting weights. Otherwise, for example, weights of the persons in families with 6 adults/children would have to be increased by a factor of six, which would add substantial variation in the weights. Note that the adjustment is made at the overall level by family categories instead of at the individual family level also for avoiding large increase in the variation of weights.

Specifically, this initial adjustment for the sample adult link weight was made for the *i*th adult (18+ years) in the MEPS file as follows:

$$wa_{meps,i} = w_{meps,i} * \frac{\sum_{i \epsilon c} w_{meps,i}}{\sum_{i \epsilon c \cap L} w_{meps,i}} \qquad \text{if } i \in L$$
$$= 0 \qquad \text{otherwise}$$

where, $wa_{meps,i}$=initial adjusted weight for the *i*th adult, $w_{meps,i}$=MEPS FY weight for the ith adult, *c=1,2,3+* represents the cells for the categories of the number of adults in the family and L represents the set of linked adults.

Similarly, this initial adjustment for the sample child link weight was made for the *i*th child (2-19 years) in the MEPS file as follows:

$$wc_{meps,i} = w_{meps,i} * \frac{\sum_{i \epsilon c} w_{meps,i}}{\sum_{i \epsilon c \cap L} w_{meps,i}} \qquad \text{if } i \in L$$
$$= 0 \qquad \text{otherwise}$$

where, $wc_{meps,i}$=initial adjusted weight for the *i*th child, $w_{meps,i}$=MEPS FY weight for the ith child, *c=1,2,3+* represents the cells for the categories of the number of children in the family and L represents the set of linked children.

Table 4. Raking dimensions used in weighting adjustments for the linked files

	Raking Dimensions	
	Full File	**Sample Adult and Sample Child Files**
1.	Panel*Census Region*Age Category7 (2-4, 5-19, 20-29, 30-44, 45-64, 65-74, 75+)*Sex	Age Category8 (2-4, 5-9, 10-17, 18-29, 30-44, 45-64, 65-74, 75+)*Sex
2.	Race Ethnicity2 (Hispanic/Black, All Other)*Census Region*Poverty Category5 (Poor, Near Poor, Low Income, Middle Income, High Income)*Sex	Age Category4 (2-17, 18-29, 30-64, 65+)* Race Ethnicity3 (Hispanic, Black, All Other)*Sex
3.	Panel*Census Region*Race Ethnicity4 (Hispanic, Black, Asian, All Other)	Age Category4* Poverty Category3 (Poor, Near Poor/Low Income, Middle /High Income)
4.	Race Ethnicity2*Poverty Category3*Agecat7	Census Region * Race Ethnicity4
5.	Race Ethnicity4*Age Category7*sex	Census Region*Age Category4*SEX
6.	Census Region*MSA (msa, non-msa)*Sex	Race Ethnicity3* Poverty Category3
7.	Race Ethnicity3*Poverty Category5*SEX	Census Region*Poverty Category5
8.	Census Region*MSA*Age Category7	Census Region*MSA*Sex
9.	Panel*MSA*Race Ethnicity3	Panel*MSA*Sex
10.	Race Ethnicity3*Census Region3*AgeCategory7	Panel*Age Category4*Sex
11.	Census Region*MSA*Poverty Category3	Panel* Race Ethnicity3*Sex
12.	Panel*Age Category7*Poverty Category3	Panel* Census Region*MSA
13.		MSA*Age Category8*Sex
14.		Age Category4*Insurance (Any Private, Public only, Uninsured)

As mentioned above, no specific adjustment is made to derive the initial weight for non-linkage for the full linked file. Therefore, the initial weight used for raking the full linked file is equal to $w_{meps,i}$=MEPS FY weight for the ith person in the full file.

The raking dimensions used for adjustments of the initial weights in the three files are shown in Table 4. The variables and dimensions used in the raking adjustments are similar to those used for non-response and raking adjustments in deriving the MEPS FY weights. The control totals used for raking adjustments are obtained by adding the MEPS FY weights for both linked and non-linked cases for the relevant subgroups. These control totals are consistent with the CPS population totals because the MEPS weight is derived by using the control totals from the CPS population totals. Table 5 presents the distributions of weights before and after weighting adjustments for the three files.

Table 5. Distribution of MEPS weight and Link weight for linking with NHIS Full, Sample Adult and Sample Child Files

Linked with	Weight	Number of cases	CV	STD	Min	First Quartile	Median	Third Quartile	Max
Full File	MEPS Weight	28,539	67.5	6,925.6	550	5,150	8,920	13,528	67,154
	Linked Weight	27,338	67.4	7,218.4	585	5,436	9,346	14,018	71,980
Sample Adult	MEPS Weight	20,873	64.2	6,993.4	566	5,800	9,640	14,200	67,154
	Linked Weight	9,167	75.5	18,738.2	822	12,359	20,772	31,521	246,940
Sample Child	MEPS Weight	8,608	74.6	6,421.8	550	3,955	6,971	11,468	60,441
	Linked Weight	3,690	71.2	14,302.2	1,097	10,041	16,642	26,242	156,549

4. Results - Comparison of Estimates

In this section, we compare the estimates for selected MEPS and NHIS variables based on the linked cases with adjusted and unadjusted weights to see if the weight adjustment for non-linking has a substantial impact on the estimates. To facilitate the comparison, another set of estimates, which will be used as the benchmark for the comparison, is produced based on the full MEPS or NHIS files (i.e. using both linked and non-linked cases) with the usual MEPS or NHIS weights. The NHIS full file is created by combining 2005 and 2006 NHIS files because the 2007 MEPS FY file includes two panels which were selected from the 2005 and 2006 NHIS households. Since the two years data are combined, a compositing factor is applied to the NHIS weights to scale down the weights to represent one year. The weighted proportions of MEPS cases linked with the 2005 and 2006 NHIS files are used to compute the compositing factor.

Similar comparison of estimates is done for the sample adult and the sample child link files. The comparison scheme for estimates from different link files and different weights are summarized in Table 6. The adjusted and unadjusted estimates from a link file are compared against the benchmark estimates to see if the adjusted estimates are substantially closer to the benchmark estimates. Statistical significance of the differences of adjusted and unadjusted estimates from the benchmark estimates are indicated in respective tables.

Table 6. Description of estimates compared

	MEPS/NHIS Full File (linked + non-linked cases)	**MEPS–NHIS linked File** (linked cases only)	
Comparison for	**Benchmark Estimates** (using MEPS/NHIS weights)	**Adjusted Estimates** (using MEPS weight adjusted for non-linkage)	**Unadjusted Estimates** (using MEPS weight not adjusted for non-linkage)
MEPS (age 2+ years) linked with NHIS full file	For MEPS variables: MEPS full file (2+ years) * and MEPS weight (PERWT07F). For NHIS variables: NHIS full file (2+ years)* with NHIS weight (WTFA).	For both MEPS and NHIS variables: MEPS-NHIS linked file and full sample Linked weight (LINKWT07)	For both MEPS and NHIS variables: MEPS-NHIS linked file and MEPS weight (PERWT07F)
MEPS (age 18+ years) linked with NHIS Sample Adult file	For MEPS variables: MEPS full file (18+ years) and MEPS weight (PERWT07F). For NHIS variables: NHIS Sample Adult file (18+ years)* with NHIS Sample Adult weight (WTFA_SA).	For both MEPS and NHIS variables: MEPS-NHIS Sample Adult linked file and sample adult linked weight (LINKWT07A)	For both MEPS and NHIS variables: MEPS-NHIS Sample Adult linked file and MEPS weight (PERWT07F)
MEPS (2-17 years) linked with NHIS Sample Child file	For MEPS variables: MEPS full file (2-17 years) and MEPS weight (PERWT07F). For NHIS variables: NHIS Sample Child file (2-17 years)* with NHIS Sample Child weight (WTFA_SC).	For both MEPS and NHIS variables: MEPS-NHIS sample child linked file and sample child linked weight (LINKWT07C)	For both MEPS and NHIS variables: MEPS-NHIS Sample Child linked file and MEPS weight (PERWT07F)

*All ages are at the MEPS poststratification/raking reference date of 12/31/07

Linking MEPS (age 2+ years) with NHIS Full Sample

For linking of MEPS with the NHIS full sample, Table 7 presents the comparison of estimates using selected MEPS variables. None of the adjusted or unadjusted estimates from the linked file are significantly different from the benchmark estimates produced from the full MEPS file. In addition to the high link rate, this may be due to the absence of any systematic non-link pattern as reflected in the analysis by background characteristics in Table 3. However, in most cases, the adjusted estimate is closer to the benchmark estimate compared to the unadjusted estimate. For example, for the mean of total healthcare expenditures, the adjusted estimate ($3,773) is closer to the benchmark estimate ($3,778) than the unadjusted estimate ($3,824). Table 8 presents the same

Table 7. Comparison of estimates using MEPS variables for linking 2007 MEPS (2+ years) with pooled 2005-06 NHIS full file

Variable	Estimates from MEPS FY File (n=28,538)		Estimates* from Linked File (n=27,338)			
	Benchmark		Adjusted		Unadjusted	
	Mean/Pct	SE	Mean/Pct	SE	Mean/Pct	SE
Total healthcare expense ($)	3,778.3	88.88	3,772.5	86.08	3,823.6	87.38
No healthcare expense (%)	15.4	0.35	15.2	0.340	15.0	0.340
Amount paid by private insurance ($)	1,571.5	64.75	1,558.9	59.90	1,582.6	61.45
Zero amount paid by private insurance (%)	43.7	0.62	43.3	0.61	42.9	0.61
Total out-of-pocket expenditures ($)	618.4	12.80	621.0	12.84	630.0	13.00
No out-of-pocket expenditure (%)	23.0	0.43	22.8	0.43	22.4	0.42
Total expenditures for office-based visits ($)	596.2	16.76	601.0	17.03	608.4	17.24
No expenditure for office-based visit (%)	34.9	0.44	34.6	0.44	34.4	0.44
Insured on 12/31/07 (%)	81.5	0.42	81.7	0.43	81.8	0.43
Private insurance on 12/31/07 (%)	63.4	0.60	63.7	0.60	64.0	0.59

*No estimate is significantly different than the benchmark estimate at ≤5% level

Table 8. Comparison of estimates using NHIS variables for linking 2007 MEPS (2+ years) with pooled 2005-06 NHIS full file

Variable	NHIS full sample 2005-2006 (n=173,188)		Estimates from Linked File (n=27,338)			
	Benchmark		Adjusted		Unadjusted	
	Est (%)	SE	Est (%)	SE	Est (%)	SE
Any limitation of activity	12.2	0.13	11.8	0.3	11.9	0.3
Medical care delayed	7.6	0.10	8.1	0.29	8.1	0.3
Needed but did not get medical care (PNMED12M)	5.5	0.09	5.8	0.25	5.8	0.25
With Medicare	13.0	0.16	12.4*	0.33	12.7	0.33
With Medicaid	9.8	0.15	9.9	0.38	9.6	0.37
With private insurance	65.2	0.27	65.8	0.62	66.1	0.61
Family spent no money for medical care	8.5	0.15	8.3	0.37	8.1	0.36
Born in the US	86.1	0.19	86.0	0.43	86.1	0.43
Has US citizenship	92.2	0.14	92.3	0.33	92.4	0.33
Own/buying house	69.0	0.33	68.8	0.75	69.2	0.75

*Significantly different than the benchmark estimate at ≤5% level

comparisons but using selected NHIS variables. Similar to MEPS variables, the adjusted and unadjusted estimates are mostly similar and reasonably close to the benchmark estimates (only 1 statistically significant difference).

Linking MEPS (18+ years) with NHIS Sample Adult

For linking with the NHIS Sample Adult, Table 9 presents the comparison of estimates for selected MEPS variables and Table 10 presents the comparison of estimates for selected NHIS variables. The differences between the three estimates are slightly higher here than the comparison for linking with the NHIS full sample. This is expected as the link rate is only 45.8% for linking with the adult sample file compared to 95.7% for linking with the NHIS full sample. However, in many cases adjusted and unadjusted estimates for MEPS variables (Table 9) are very close, but when there is a difference, the adjusted estimates are generally closer to the benchmark estimates than the unadjusted estimates. For example, the estimate of mean total healthcare expenditures obtained from the linked file with the adjusted weight ($4,611) is closer to the benchmark estimate ($4,501) compared to the unadjusted estimate ($5,112). For most other MEPS variables compared such as personal income, out-of-pocket expenditures, OBD expenditures, and percent insured, the adjusted estimates are also closer to the benchmark estimates compared to the unadjusted estimates.

Table 9. Comparison of estimates using MEPS variables for linking 2007 MEPS (18+ years) with pooled 2005-06 NHIS sample adult file

Variable	Estimates from MEPS FY File (n=20,873)		Estimates from Linked File (n=9,167)			
	Benchmark		Adjusted		Unadjusted	
	Mean/Pct	SE	Mean/Pct	SE	Mean/Pct	SE
Total healthcare expense ($)	4,501.4	105.88	4,611.9	135.10	5,112.0*	140.14
No healthcare expense (%)	15.4	0.36	14.1*	0.51	12.0*	0.41
Amount paid by private insurance ($)	1,860.8	82.26	1,802.7	88.53	1,877.6	88.76
Zero amount paid by private insurance (%)	42.2	0.57	41.3*	0.72	40.3*	0.67
Total out-of-pocket expenditures ($)	721.7	15.06	763.9*	25.24	834.2*	25.95
No out-of-pocket expenditure (%)	19.5	0.37	18.1*	0.57	15.6*	0.48
Total expenditures for office-based visits ($)	701.3	20.89	714.8	35.10	759.8*	26.81
No expenditure for office-based visit (%)	34.2	0.45	32.4*	0.64	29.3*	0.59
Insured on 12/31/07 (%)	79.7	0.43	80.1	0.56	82.1*	0.48
Private insurance on 12/31/07 (%)	64.9	0.54	65.2	0.67	65.0	0.63

*Significantly different than the benchmark estimate at ≤5% level

The pattern is fairly similar when estimates of NHIS variables are compared with benchmark estimates obtained from the NHIS full file. For example, the percent with 'delayed medical care for costs', the adjusted estimate (10.4%) is not significantly different from the benchmark estimate (9.7%) whereas the unadjusted estimate (11.4%) is

significantly different; or for the percent 'who needed but did not get medical care' – the adjusted estimate (7.7%) is not significantly different from the benchmark estimate (7.2%), whereas the unadjusted estimate (8.3%) is significantly different. In some other cases, while both the adjusted and the unadjusted estimates are significantly different than the benchmark estimate, the adjusted estimate is usually closer to the benchmark estimate. For example, with respect to the percent with 'any limitation of activity', the adjusted estimate (13.9%) is considerably closer to the benchmark estimate (14.8%) than the unadjusted estimate (16.2%) to the benchmark estimate.

Table 10. Comparison of estimates using NHIS variables for linking 2007 MEPS (18+ years) with pooled 2005-06 NHIS sample adult file

Variable	NHIS Sample Adult 2005-2006 (n=55,701)		Estimates from Linked File (n=9,167)			
	Benchmark		Adjusted		Unadjusted	
	Est (%)	SE	Est (%)	SE	Est (%)	SE
Any limitation of activity	14.8	0.19	13.9*	0.44	16.2*	0.49
Medical care delayed	9.7	0.16	10.4	0.44	11.4*	0.47
Needed but did not get medical care	7.2	0.13	7.7	0.39	8.3*	0.41
With Medicare	17.6	0.22	16.2*	0.48	20.1*	0.53
With Medicaid	6.6	0.13	6.9	0.32	7.1	0.31
With private insurance	66.8	0.3	66.7	0.69	66.6	0.67
Family spent no money for medical care	8.9	0.16	9.9*	0.44	10.1*	0.45
Born in the US	84.2	0.25	83.4	0.57	85.7*	0.5
Has US citizenship	91.7	0.18	91.4	0.41	92.8*	0.36
Own/buying house	70.1	0.36	65.5*	0.9	64.5*	0.89
Has cancer	7.3	0.13	7.0	0.33	7.9	0.35
Has diabetes	7.6	0.14	7.2	0.3	7.8	0.31

*Significantly different than the benchmark estimate at ≤5% level

Linking MEPS (2-17 years) with NHIS Sample Child

For linking with the NHIS Sample Child, Tables 11 and 12 present the comparison of estimates for selected MEPS and NHIS variables respectively. For MEPS variables (Table 11), the adjusted estimates are generally closer to the benchmark estimates than the unadjusted estimates. More of the unadjusted estimates are significantly different from the benchmark estimates. For example, the adjusted estimate of 57.9% with private insurance is not significantly different from the benchmark estimate of 58.1%, but the unadjusted estimate of 61.6% is significantly different from the benchmark. For mean total out-of-pocket expenditures, again the adjusted estimate of $277 is not significantly different from the benchmark estimate of $260, but the unadjusted estimate of $306 is significantly different from the benchmark estimate. For most of the estimates compared, irrespective of whether the differences of both adjusted and unadjusted estimates from the benchmark estimate are statistically significant or not, the adjusted estimates are closer to the benchmark estimates. For example, for mean total healthcare expenditures, the adjusted estimate of $1,334 is much closer to the benchmark

Table 11. Comparison of estimates using MEPS variables for linking 2007 MEPS (2-17 years) with pooled 2005-06 NHIS sample child file

Variable	Estimates from MEPS FY File (n=7,665)		Estimates from Link File (n=3,263)			
	Benchmark		Adjusted		Unadjusted	
	Mean/Pct	SE	Mean/Pct	SE	Mean/Pct	SE
Total healthcare expense ($)	1270.8	116.04	1334.2	157.11	1517.4	243.49
No healthcare expense (%)	15.3	0.62	14.4*	0.73	13.2*	0.68
Amount paid by private insurance ($)	568.1	37.45	568.7	46.59	642.9	59.74
Zero amount paid by private insurance (%)	48.7	1.14	48.4	1.22	45.0*	1.15
Total out-of-pocket expenditures ($)	260.3	14.14	276.8	21.57	305.8*	25.33
No out-of-pocket expenditure (%)	35.0	0.96	34.2	1.07	30.7*	1.04
Total expenditures for office-based visits ($)	232.0	8.69	240.6	12.02	259.2*	14.66
No expenditure for office-based visit (%)	37.2	0.83	35.5*	1.00	33.6*	1.00
Insured on 12/31/07 (%)	87.8	0.69	87.3	0.78	87.8	0.77
Private insurance on 12/31/07 (%)	58.1	1.15	57.9	1.17	61.6*	1.09

*Significantly different than the benchmark estimate at ≤5% level

Table 12. Comparison of estimates using NHIS variables for linking 2007 MEPS (2-17 years) with pooled 2005-06 NHIS sample child file

Variable	NHIS Sample Child 2005-2006 (n=19,154)		Estimates from Linked File (n=3,263)			
	Benchmark		Adjusted		Unadjusted	
	Est (%)	SE	Est (%)	SE	Est (%)	SE
Any limitation of activity	7.5	0.24	7.8	0.59	7.7	0.59
With private insurance	59.2	0.51	59.7	1.14	63.3*	1.07
Family spent no money for medical care	9.7	0.26	9.1	0.62	8.5*	0.62
Has US citizenship	96.5	0.15	96.8	0.32	97.2*	0.29
Own/buying house	66.8	0.49	67.3	1.08	68.7	1.09
No. of times in ER/ED	78.3	0.39	77.9	0.87	77.6	0.86
Seen/talked to a general doctor in past 12 month	80.2	0.38	80.2	0.85	80.6	0.87
Had well-child checkup	73.1	0.44	71.9	0.98	72.6	0.97
Ever been told had Asthma	12.7	0.29	12.3	0.7	12.4	0.69

*Significantly different than the benchmark estimate at ≤5% level

estimate of $1,271 than the unadjusted estimate of $1,517. For amount paid by private insurance, the adjusted estimate of $569 is virtually identical to the benchmark estimate of $568, whereas the unadjusted estimate of $643 is significantly higher.

For NHIS variables (Table 12), again the unadjusted estimates are more often significantly different from the benchmark estimates. For example, the adjusted estimate of 59.7% for having private insurance is not significantly different from the benchmark estimate of 59.2% whereas the unadjusted estimate of 63.3% is significantly different.

5. Conclusion

In this study, we evaluated the impact of weighting adjustment when analyzing a MEPS-NHIS linked dataset. We focused the analysis on the situation where one or more NHIS variables are added to enhance the analytic capability of the MEPS. The MEPS weight is adjusted to account for the non-linked cases when linking with three NHIS files – full, sample adult, and sample child files. The estimates obtained using unadjusted and adjusted weights from each linked file are compared to the corresponding benchmark estimates obtained from the full MEPS and the full NHIS files with regular weights.

The analysis shows that for linking with the NHIS full file, the weighting adjustment for non-linkage has a negligible impact on the estimates. The unadjusted estimates are reasonably close to the adjusted estimates in most cases, which would be expected since the non-linkage rate is very low. However, even though differences are not significant, the adjusted estimates are closer to the benchmark estimates than the unadjusted estimates in most cases.

For linking with NHIS sample adult or sample child files, where the non-linkage rate is very high, even though adjusted and unadjusted estimates are often similar, the differences between the estimates are noticeable in many cases. In those cases, the adjusted estimates are generally closer to the benchmark estimates and the unadjusted estimates are more often significantly different from the benchmark estimates. Therefore, it is recommended to adjust the weights for MEPS-NHIS linked analysis when linking with the sample adult or sample child files.

For linking with the NHIS full file, where the non-linkage rate is very low, the risk of bias due to non-linkage is negligible and the analysis can be done using a simple overall adjustment. In some cases, the analysis can also be done by considering the relevant NHIS variables as missing for the non-linked cases. An alternative option is to impute the missing values of the relatively small number of non-linked cases. This option warrants further investigation.

References

Bernard, D.S.M., Farr, S.L., Fang, Z. 2011. National Estimates of out-of-pocket health care expenditure burdens among nonelderly adults with cancer: 2001 to 2008. Journal of Clinical Oncology, July 10, 2011.

Cohen, S. 2010. An estimation methodology to permit longitudinal analysis based on the National Health Interview survey and Medical Expenditure Panel Survey. Agency for Healthcare Research and Quality, Working Paper No. 11002. http://meps.ahrq.gov/mepsweb/data_files/publications/workingpapers/wp_11002.pdf

Machlin S.R., Chowdhury S.R., Ezzati-Rice T., DiGaetano R., Goksel H., Wun L.-M., Yu W., Kashihara D. 2010. Estimation Procedures for the Medical Expenditure Panel Survey Household Component. Methodology Report #24. September 2010. Agency for Healthcare Research and Quality, Rockville, MD. http://www.meps.ahrq.gov/mepsweb/data_files/publications/mr24/mr24.pdf

Rhoades, J., and Cohen, S. 2012. Statistical Brief #335: The Long-Term Uninsured in America, 2007–2010 (Selected Intervals): Estimates for the U.S. Civilian Noninstitutionalized Population under Age 65. August, 2012. Agency for Healthcare Research and Quality, Rockville, MD. http://meps.ahrq.gov/mepsweb/data_files/publications/st335/stat335.pdf