

## Accuracy of Household-Reported Expenditure Data in the Medical Expenditure Panel Survey

David Kashihara<sup>1</sup> and Diana Wobus<sup>2</sup>

<sup>1</sup>Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850

<sup>2</sup>WESTAT, 1650 Research Boulevard, Rockville, MD 20850

**Key Words:** medical reports, response accuracy, survey quality

### Introduction

Due to the nature of medical expense payments, it is often difficult for respondents in household surveys to know exactly how much was paid to medical providers for services they received. Sometimes, payments for medical events include a payment from a third-party payer (e.g. private insurance). This can be accompanied by a co-payment from the insured person. Other times the third-party may pay for the entire medical event. When the person is uninsured, that person may have paid for the entire event themselves. Many different scenarios exist and each one may have an effect on how complete and how accurately the household respondent is able to report medical expenditures when participating in a survey. Because the accuracy of household reports of medical expenditures directly affects the quality of the estimates of medical expenditures, it is very important to obtain accurate medical expense reports from sample persons that participate in household surveys.

In this paper, the “accuracy” of a payment report indicates how accurately the household reported the payment information, given that a complete report was provided. This study uses data from both the Household Component (HC) and the Medical Provider Component (MPC) of the 2003 Medical Expenditure Panel Survey (MEPS) to assess the accuracy of complete medical payment reports. Accuracy is examined by respondent, household, and other characteristics by three types of payment source categories: out-of-pocket, private insurance, and the total of all payment sources. The ten potential sources of payments are:

1. family/patient out-of-pocket
2. Medicare
3. Medicaid
4. private insurance
5. Veterans Administration
6. TRICARE
7. other Federal (e.g., Indian Health Service, military treatment facilities, federally funded NIH care)
8. other State (e.g., community/neighborhood clinics, state and local health departments, state programs other than Medicaid)

9. worker’s compensation
10. other sources (e.g., automobile, homeowner’s, or liability insurance payments, miscellaneous)

A particular medical event may involve just one source of payment. However, it is not uncommon for several different sources of payments to be associated with the same event.

The MEPS-HC is a large, national probability sample survey sponsored by the Agency for Healthcare Research and Quality. The annual survey collects information from respondents to produce national- and regional-level estimates of health care use, health status, health conditions, medical expenditures, sources of payment, insurance coverage, and health care access for the U.S. civilian noninstitutionalized population as well as for policy-relevant sub-groups. The MEPS-HC is a two year panel survey with a new panel introduced each year. Five rounds of interviews are conducted with each new panel to yield health care use and expenditure data for two calendar years. The MEPS sample is a sub-sample of respondents to the prior year’s National Health Interview Survey that is conducted by the National Center for Health Statistics (Cohen, 2000).

The MEPS-MPC is a telephone survey of a sample of providers that were identified in the household survey. The MPC is conducted the year following the household data collection year and is dependent on receiving permission from the respondent to contact the medical provider and obtaining cooperation from the provider to participate in the survey. Because the MPC is based on records from the medical provider, it is generally more complete and more accurate than information provided by household respondents.

This study compares the 2003 HC data and the 2003 MPC data to examine the accuracy of household reports of payments for office-based physician visits. These visits will be referred to as “events” in this paper. Events that are matched between the HC and the MPC and are classified as being complete on both sources are used for the analysis. Multivariate

logistic models are run to build a profile of those who provide accurate reports.

**Methods**

To determine the accuracy of medical payment reports, we used matched and complete records from the 2003 HC and the 2003 MPC. For office-based physician visits (events) in the 2003 HC survey, the number of permission forms obtained varied by survey round but was about ¾ of the households. MEPS sampled about 40% of the medical providers and their cooperation rate was 83.5%. Once the MPC reports were obtained, each event was matched to reports from the HC using the probabilistic matching software AUTOMATCH, which utilizes the Fellegi and Sunter methodology (Winglee et al, 2000). In 2003, the match rate was 83.8% overall.

For both the HC and the MPC data, for events to be classified as being “complete”, the events must include payment information from all requisite payment sources. The combination of payment sources can vary widely for each individual event.

This study of the accuracy of household medical reports concentrates on three types of payment sources: Total of All Sources, Out-of-Pocket, and Private Insurance. The number of complete and matched records for each payment source for office-based physician visits is shown in Table 1. Note that flat fee records, where one fee was paid for multiple office visits, were not included in this analysis.

**Table 1** – Number of Complete and Matched Records by Source of Payment Categories, Office-based Physician Visits, 2003 MEPS

<u>Payment Source</u>	<u>Sample Size</u>
Total of all sources	6,106
Out-of-Pocket	43,128
Private Insurance	4,708

Thus, for each payment source category, an event will qualify for this study if it is complete for that payment source category – even if it is missing other payment sources. For example, if we look at an event that involves an out-of-pocket payment, if the event is complete for the out-of-pocket payment source but is not complete for the private insurance payment source, the event is still eligible in this study for the out-of-pocket analysis. However, that same event would not be eligible for the total-of-all-sources analysis.

It is assumed that provider reports of medical expenditures are generally more accurate than household reports because the providers know the payment amounts and from which sources they originated. However, the provider reports are not

perfect. Errors from abstracting and recording the data may occur. Also, the matching process between household-reported and provider-reported events uses AUTOMATCH, a probabilistic matching software. Thus, since the process is probabilistic, there may be some false-negative as well as some false-positive matches.

The criteria used to determine the accuracy of the household reports when compared to the provider reports were similar to that used in previous research performed on the 1996 MEPS data (Machlin et al, 1999) and generally reflect the relative sizes of the mean payments for the various sources (Table 2).

**Table 2** – Mean Expenses by Payment Source Category in the HC and the MPC, Office-based Physician Visits, 2003 MEPS

<u>Payment Source</u>	<u>Mean Expense</u>	
	<u>HC</u>	<u>MPC</u>
	<u>(s.e.)</u>	<u>(s.e.)</u>
Total of all sources	\$157 (8.8)	\$129 (5.7)
Out-of-Pocket	\$ 15 (1.3)	\$ 12 (0.5)
Private Insurance	\$ 79 (5.6)	\$ 78 (5.2)

For the Total of all Sources payment source category, if the HC payment value was within \$20 or 10% of the MPC value, the report was deemed accurate. For example, if a person had a \$200 total expense and they reported between \$180 and \$220, they would be considered to have provided an accurate report. For the Out-of-Pocket source, if the HC value was within \$5 or 10% of the MPC value, the report was deemed accurate. For the Private Insurance payment source, if the HC value was within \$10 or 10% of the MPC value, the payment was deemed accurate.

To examine the variation in accuracy, multivariate analyses were performed on each of the three payment sources. Logistic regression models were developed with the dependent variable being “accurate” (1) or “not accurate” (0). The explanatory variables represented a wide variety of factors that may influence the accuracy of household medical expenditure reports.

In MEPS, one person in the household responds for everyone in that household. Thus, sometimes the actual respondent may be different than the survey person about whom the report concerns. Proxy respondents may be neighbors or friends who are not part of the household. Therefore, because there is not any characteristic information about these proxy respondents and there are not many of them, events reported by them were excluded from this study.

We grouped explanatory variables as follows: respondent characteristics, household characteristics, person insurance coverage, office-visit characteristics, and survey procedure characteristics. More specifically, the respondent characteristics used were age, gender, health status, race/ethnicity, education, and respondent type (self or family member). The household characteristics used were family poverty status, region, and MSA status. We suspected that the type of insurance (or lack of insurance) would have an impact on how accurately they would be able to report their expenses. For example, those in an HMO may not know exactly how much was paid to the provider for a particular event or the provider may not be paid on a per-event basis, whereas someone who was uninsured may have a better chance of providing an accurate report. The characteristics for office-visits were the amount of payments as reported by the medical provider and the number of medical events in a round. The amount of payments variable was included in the model under the hypothesis that those persons with larger payments would be less likely to provide an accurate expense report (as defined in this study) of their medical events. The number of medical events in a round was included as a measure of burden on the respondent – a higher burden may cause a loss of accuracy due to the greater number of events. For survey procedure characteristics, the round variable was introduced into the model to control for variation in accuracy between the five rounds of data collection. Finally, the reporting aids variable would indicate what materials, if any, the respondent had to accurately report their expenses.

The reporting aids variable was constructed in a hierarchical fashion. First, use of an explanation of benefits form from an insurer along with any other aids was classified as “explanation of benefits”. Next, use of any bill from a medical provider along with any other aids was classified as “any bill”. After that, use of the MEPS monthly planner, provided by the interviewer in a previous visit, along with any other aids was classified as “any calendar”. Then, the use of a checkbook, pill bottle, or other aid was classified as “any checkbook/bottle/other”. Finally, if the respondent did not use any aids, the variable was classified as “memory only”.

**Results**

The accuracy criteria used in this study are summarized in Table 3. To determine if a value reported by the household on the HC was “accurate,” it would have to fall within these constraints when compared to the matched value reported by the medical provider on the MPC.

**Table 3** – Accuracy Criteria for Office-based Physician Visits by Source of Payment Categories, 2003 MEPS

<u>Payment Source</u>	<u>Accuracy Criteria</u>
Total of All Sources	\$20 or 10%
Out-of-Pocket	\$ 5 or 10%
Private Insurance	\$10 or 10%

Based on these accuracy criteria, the percent accurate varied for the three payment source categories as shown in Table 4.

**Table 4** – Accuracy Rates for Office-based Physician Visits by Source of Payment Categories, 2003 MEPS

<u>Payment Source</u>	<u>Percent Accurate</u>
Total of All Sources	50.0
Out-of-Pocket	77.9
Private Insurance	50.2

A multivariate logistic regression model for reporting accuracy was developed for each type of payment source. Table 5 (Appendix) shows the logistic regression coefficients and the odds ratios (OR) of the explanatory variables in each of the three models.

At the 0.05 level of significance, the significant variables in the total-of-all-sources model were the amount of MPC-reported payments, reporting aids, type of insurance, health status, and race/ethnicity. As seen in Table 5, payments greater than \$300 were less likely to be reported accurately than payments of zero dollars (OR = 0.20). With reporting aids, use of an explanation of benefit form led to more accurate reporting than memory only (OR = 2.70). Use of a bill also led to more accurate reporting than memory (OR = 1.89). The non-elderly (under 65 years of age) uninsured reports were more likely to be accurate when compared to any private insurance HMO plan members (OR = 1.77). Also for the non-elderly, any private insurance fee for service plan members were more likely to provide accurate reports than any private insurance HMO plan members (OR = 1.34). Persons in better health had more accurate reports than those in fair or poor condition (OR = 1.58). And Hispanics were less likely to accurately report than Whites (OR = 0.66).

For the out-of-pocket model, the significant variables at the 0.05 level of significance were the amount of MPC-reported payments, type of insurance, number of events, and region. At the 0.10 level of significance, family poverty status and reporting aids also were significant. Table 5 shows that those with payment amounts greater than zero dollars were less likely to report accurately (OR = 0.20 for \$1-\$50, 0.07 for \$51-\$100, and 0.05 for >\$100). Non-elderly persons with any private fee for service plan had less accurate reports than those with any private HMO

plan (OR = 0.70). Also, non-elderly persons with public only insurance had more accurate reports than those with any private HMO (OR = 2.60). Those with two to five events were less likely to have accurate reports than those with just one event (OR = 0.84 for 2 events and 0.84 for 3 to 5 events). Residents in the Midwest were less accurate reporters than those in the northeast (OR = 0.77). Those with low family income were less likely to provide an accurate report than the poor (OR = 0.77) and those with middle family income were also less likely to provide accurate reports than the poor (OR = 0.78). For reporting aids, those who used the MEPS monthly planner provided more accurate reports than those who relied on memory alone (OR = 1.20).

For the private insurance source of payment model, at the 0.05 level of significance, the significant variables were the amount of MPC-reported payments, type of insurance, reporting aids, race/ethnicity, and age. Table 5 shows that those with higher payment amounts were less likely to report accurately than those with zero payment amounts (OR = 0.74 for \$1-\$100, 0.33 for \$101-\$200, and 0.13 for > \$200). Elderly persons with Medicare and private insurance or Medicare and Medicaid were more likely to have accurate reports than non-elderly people with any private insurance, HMO, or fee for service insurance (OR = 2.42). Use of reporting aids led to increased accuracy than just memory only (OR = 2.90 for any explanation of benefit, 1.59 for any bill, 1.54 for any calendar, and 1.94 for the use of a checkbook/bottle/other aid). When examining race/ethnicity, Blacks were less likely to report accurately than Whites (OR = 0.58) and Hispanics were also less likely to report accurately than Whites (OR = 0.58). The elderly, those 65 years or older, provided less accurate reports than those aged 25 to 64 (OR = 0.47).

## Discussion

The process to develop the expenditure estimates in MEPS is lengthy. First, medical events for all family members as well as their providers must be identified by the household respondent. Once identified, the respondent must also give permission for MEPS to contact the provider. The provider is then contacted and information from the respondent's medical records is obtained. The data from the provider are then matched to the data from the household respondent wherever possible. When matched, the data from the provider are used because the providers generally have greater knowledge of the amounts of payments from the various possible sources. The matching process for linking the household-reported and provider-reported medical events uses a probabilistic matching method. When no provider data are available, the household data are used. Any missing information is imputed. Thus, it is important

to obtain accurate household reports of medical expenditures because some of those data are used in the estimation process. In 2003, MEPS expenditure data for nearly one in five physician office visits are based on household reports. Table 6 (Appendix) summarizes the factors associated with accuracy for the three payment sources examined in this study.

The factors that were significant in all three models were: amount of medical provider-reported payments, type of insurance, and the use of reporting aids. To help respondents report expenses more accurately, it has been shown here that reporting aids play a very important role. However to improve the quality of the survey expenditure estimates, different sub-populations should also be monitored closely. For example, people with a greater likelihood of high expenditures and also with certain types of insurance could be targeted for the greatest level of assistance to improve the accuracy of their reports. From the Total-of-all-Sources model, we see that those in fair/poor health might be able to use additional assistance to improve the quality of their reports. From the Out-of-Pocket model, we see that those with more events, higher income, and those living in the Midwest might benefit from additional methods to improve the accuracy of their reports. If improved accuracy is obtained in the household medical reports, higher quality estimates of expenditures will result.

Other areas of related research could be conducted to obtain a better understanding of the nature of the accuracy of household medical reports. For example, the direction of misreporting would provide a glimpse at a possible source of bias in the estimates. In an analysis of 1996 MEPS data, Machlin et al discovered an over-reporting trend of household-reported payments (Machlin et al, 1999). Also, other sources of payments could be examined to see how they are impacted by each of the explanatory variables used in this study. As seen in this analysis, the three types of payment sources differed slightly in the number and the types of significant variables.

In summary, the availability of medical provider records greatly improves the accuracy of the expenditure estimates in the MEPS survey. However, the MPC is only a sample of medical providers – not every provider is contacted. Therefore, it is important to maintain and, if possible, improve the accuracy of the household reports so that the medical expenditure estimates produced from the survey will be of the highest possible quality.

## Acknowledgements

The authors would like to thank Steven Machlin for his invaluable contributions to this study.

## References

Cohen S.B. (2000), "Sample Design of the 1997 Medical Expenditure Panel Survey Household Component". MEPS Methodology Report No. 11. AHRQ Pub. No. 01-0001. Rockville, MD: Agency for Healthcare Research and Quality.

Machlin, S., Cohen, J., Zuvekas, S., and Thorpe, J. (1999), "Accuracy of Household Reported Payments for Physician Visits in the 1996 Medical Expenditure Panel Survey". Proceedings of the American Statistical Association Conference, Alexandria, VA: American Statistical Association.

Machlin, S., Wobus, D., and Kashihara, D. (2006), "Completeness of Household Reported Expenditure Data in the Medical Expenditure Panel Survey (MEPS)". Slide presentation for the 2006 Joint Statistical Meetings, Seattle, WA.

Winglee, M., Valliant, M.I., Brick, J.M., and Machlin, S. (2000), "Probability Matching of Medical Events". *Journal of Economic and Social Measurement*, 2000;26(2):129-140.

---

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.

Appendices

**Table 5** – Logistic Regression Analysis: Characteristics Associated with Accuracy for the Three Types of Payment Sources for Office-based Physician Visits, MEPS 2003

(R-Square)	Total of All Sources (0.116)			Out-of-Pocket (0.157)			Private Insurance (0.091)		
Measure	Odds Ratio	Beta	S.E.	Odds Ratio	Beta	S.E.	Odds Ratio	Beta	S.E.
<b>Intercept</b>	0.73	-0.32	0.42	15.62	2.75	0.21	1.83	0.60	0.43
<b>Age</b>							*		
0 - 24	0.76	-0.28	0.27	0.93	-0.07	0.13	1.02	0.02	0.30
25 – 64	(ref)	0.00		(ref)	0.00		(ref)	0.00	
65 +	1.21	0.19	0.22	1.14	0.16	0.16	0.47	-0.76	0.27
<b>Gender</b>									
Male	0.97	-0.03	0.13	1.01	0.01	0.07	1.03	0.28	0.13
Female	(ref)	0.00		(ref)	0.00		(ref)	0.00	
<b>Health status</b>							*		
Fair/poor	(ref)	0.00		(ref)	0.00		(ref)	0.00	
Other	1.58	0.46	0.16	1.03	0.03	0.08	1.00	0.00	0.20
<b>Race/ethnicity</b>							*		
White	(ref)	0.00		(ref)	0.00		(ref)	0.00	
Black	0.72	-0.33	0.19	0.92	-0.08	0.08	0.58	-0.55	0.23
Asian/other/multiple race	1.14	0.13	0.34	1.23	0.25	0.14	0.69	-0.37	0.33
Hispanic	0.66	-0.42	0.18	1.14	0.13	0.10	0.58	-0.55	0.23
<b>Education</b>									
< High school	(ref)	0.00		(ref)	0.00		(ref)	0.00	
HS/GED/unknown	0.70	-0.36	0.23	0.90	-0.11	0.12	0.70	-0.35	0.27
College +	0.82	-0.20	0.23	0.96	-0.05	0.14	0.78	-0.24	0.27
<b>Poverty status</b>				**					
Poor/near poor/missing	(ref)	0.00		(ref)	0.00		(ref)	0.00	
Low	0.89	-0.11	0.19	0.77	-0.26	0.12	0.68	-0.38	0.32
Middle	0.86	-0.13	0.17	0.78	-0.24	0.11	0.84	-0.17	0.28
High	0.92	-0.08	0.16	0.88	-0.13	0.12	0.91	-0.10	0.27
<b>Region</b>				*					
Northeast	(ref)	0.00		(ref)	0.00		(ref)	0.00	
Midwest	1.03	0.03	0.22	0.77	-0.26	0.11	1.21	0.19	0.18
South	1.02	0.02	0.22	0.97	-0.03	0.10	0.98	-0.02	0.18
West	0.84	-0.17	0.24	0.88	-0.13	0.13	0.86	-0.15	0.21
<b>MSA status</b>									
MSA	(ref)	0.00		(ref)	0.00		(ref)	0.00	
Non-MSA	1.08	0.08	0.12	0.93	-0.07	0.07	1.02	0.02	0.12
<b>Round</b>									
1	(ref)	0.00		(ref)	0.00		(ref)	0.00	
2	1.03	0.03	0.16	1.05	0.05	0.07	0.86	-0.15	0.19
3	0.99	-0.01	0.14	1.00	0.00	0.07	1.04	0.04	0.16
4	0.90	-0.10	0.15	0.98	-0.02	0.09	0.80	-0.22	0.18
5	0.97	-0.03	0.16	0.99	-0.01	0.09	0.78	-0.24	0.19

Table 5 – continued.

Measure	Total of All Sources			Out-of-Pocket			Private Insurance		
	Odds Ratio	Beta	S.E.	Odds Ratio	Beta	S.E.	Odds Ratio	Beta	S.E.
<b>Reporting aids</b>	*			**			*		
Any EOB	2.70	0.99	0.13	1.12	0.11	0.11	2.90	1.07	0.14
Any Bill	1.89	0.64	0.10	1.16	0.15	0.09	1.59	0.47	0.13
Any Calendar	1.25	0.23	0.19	1.20	0.18	0.07	1.54	0.43	0.21
Checkbook/bottle/other	1.55	0.44	0.30	1.24	0.22	0.14	1.94	0.66	0.29
Memory only	(ref)	0.00		(ref)	0.00		(ref)	0.00	
<b>Type of Insurance</b>	*			*			*		
< 65 any private HMO	(ref)	0.00		(ref)	0.00				
< 65 any private FFS	1.34	0.30	0.12	0.70	-0.36	0.09			
< 65 public only HMO/FFS	0.81	-0.21	0.30	2.60	0.96	0.15			
< 65 uninsured	1.77	0.57	0.19	0.80	-0.22	0.14			
65 + Medicare only	0.70	-0.36	0.27	0.78	-0.24	0.21			
65 + Medicare/private & Medicare/Medicaid	0.94	-0.06	0.24	0.70	-0.35	0.17			
< 65 any private HMO/FFS							(ref)	0.00	
65 + Medicare/private or Medicare/Medicaid							2.42	0.88	0.28
<b>Respondent type</b>									
Self	0.87	-0.15	0.10	0.91	-0.09	0.06	0.94	-0.06	0.10
Family	(ref)	0.00		(ref)	0.00		(ref)	0.00	
<b>Number of events</b>				*					
1	(ref)	0.00		(ref)	0.00		(ref)	0.00	
2	0.75	-0.29	0.13	0.84	-0.17	0.06	0.86	-0.15	0.14
3 – 5	0.87	-0.14	0.11	0.84	-0.17	0.06	0.96	-0.04	0.12
6 +	0.77	-0.26	0.15	0.88	-0.12	0.08	0.93	-0.07	0.14
<b>Amount of payments</b>	*			*			*		
\$0	(ref)	0.00							
\$1 - 100	1.51	0.41	0.27						
\$101 - 300	0.60	-0.51	0.29						
> \$300	0.20	-1.64	0.38						
\$0				(ref)	0.00				
\$1 – 50				0.20	-1.59	0.08			
\$51 – 100				0.07	-2.62	0.15			
> \$100				0.05	-3.07	0.18			
\$0							(ref)	0.00	
\$1 – 100							0.74	-0.30	0.12
\$101 – 200							0.33	-1.10	0.18
> \$200							0.13	-2.02	0.20

(ref) = reference group

Wald F: \* significant at the 0.05 level, \*\* significant at the 0.10 level

Probability modeled is accurate = 1, not accurate = 0.

Sources of data: Agency for Healthcare Research and Quality, 2003 Medical Expenditure Panel Survey – Household and Medical Provider Components.

**Table 6** – Factors Associated with the Accuracy of Office-based Physician Visits, 2003 MEPS

<b>Significant Factors</b>	<b>Payment Source</b>		
	<b>All Sources</b>	<b>Out-of-Pocket</b>	<b>Private Insurance</b>
Amount of payments	√	√	√
Insurance type	√	√	√
Reporting aids	√	√	√
Race/ethnicity	√		√
Age			√
Health status	√		
Number of events		√	
Family poverty status		√	
Region		√	