

Framework for Updating DU Level Nonresponse Adjustment Covariates for the Medical Expenditure Panel Survey (MEPS)

Lap-Ming Wun and Trena M. Ezzati-Rice

Agency for Healthcare Research and Quality Working Paper No. 12001

January 2012

Suggested citation: Wun L, Ezzati-Rice TM. Framework for Updating DU Level Nonresponse Adjustment Covariates for the Medical Expenditure Panel Survey (MEPS). Agency for Healthcare Research and Quality Working Paper No. 12001, January 2012, <http://gold.ahrq.gov>.

AHRQ Working Papers provide preliminary analysis of substantive, technical, and methodological issues. The papers have not undergone external peer review. They are distributed to share valuable experience and research. Comments are welcome and should be directed to the authors. The views expressed are those of the authors and no official endorsement by the Agency for Healthcare Research and Quality or the Department of Health and Human Services is intended or should be inferred.

Framework for Updating DU Level Nonresponse Adjustment Covariates for the Medical Expenditure Panel Survey (MEPS)

By Lap-Ming Wun, Trena Ezzati-Rice

ABSTRACT

The Medical Expenditure Panel Survey (MEPS) is a complex national probability sample survey sponsored by the Agency for Healthcare Research and Quality (AHRQ). The MEPS sample is drawn from households that participated in the previous year's National Health Interview Survey (NHIS) conducted by the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention (CDC). The MEPS is designed to provide nationally representative estimates of health care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian noninstitutionalized population. Each new MEPS sample is referred to as a panel and data for each panel are collected through a series of five rounds of computer-assisted personal interviews (CAPI) that yield annual data for each of two consecutive calendar years.

MEPS, like any other large survey, experiences nonresponse. Since MEPS is a panel survey, each panel goes through five rounds of interviews, experiencing nonresponse at each of the five rounds. Procedures for making adjustments to compensate for nonresponse bias in the estimation of population parameters have been implemented in the data processing.

Since each annual MEPS sample is a subsample of respondents to the previous year's NHIS, a wide range of survey variables for all the MEPS sampled units (respondents and nonrespondents) are available from the MEPS sampling frame, the NHIS. A set NHIS variables were identified as relevant to response propensity and have been used as potential covariates for the first round dwelling unit (DU) level nonresponse adjustment. Each year, a subset of these variables is selected through CHAID (Chi-square Automatic Interaction Detection program) and used for the DU level nonresponse adjustment to the weights for that year's panel.

While adjustment variables that are optimal for reducing nonresponse bias should be related to both response propensity and survey outcome variables, the later criterion has not always been taken into consideration when selecting nonresponse adjustment covariates (Kreuter et. al. (2010)) In addition, survey paradata (i.e., information about the survey data collection process) are becoming more widely available and applied to reduce measurement error and biases in survey data. This study incorporates these two considerations as part of a review of the nonresponse adjustment covariates currently used as candidates for the MEPS round 1 DU level nonresponse adjustment.

Lap-Ming Wun
Center for Financing, Access and Cost Trends (CFACT)
Agency for Healthcare Research and Quality
540 Gaither Road
Rockville, MD 20850
Phone: (301)427-1670
Email: Lap-Ming.Wun@ahrq.gov

Framework for Updating DU Level Nonresponse Adjustment Covariates for the Medical Expenditure Panel Survey (MEPS)¹

By Lap-Ming Wun, Trena Ezzati-Rice

1. INTRODUCTION

The Medical Expenditure Panel Survey (MEPS) is a complex national probability sample survey sponsored by the Agency for Healthcare Research and Quality (AHRQ). The MEPS sample is drawn from households that participated in the previous year's National Health Interview Survey (NHIS) conducted by the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention (CDC). The MEPS is designed to provide nationally representative estimates of health care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian noninstitutionalized population. Each new MEPS sample is referred to as a panel and data for each panel are collected through a series of five rounds of computer-assisted personal interviews (CAPI) that yield annual data for each of two consecutive calendar years. Details of the MEPS sample design can be found at http://www.meeps.ahrq.gov/meepsweb/data_files/publications/mr22/mr22.pdf

MEPS, like any other large survey, experiences nonresponse. Since MEPS is a panel survey, each panel goes through five rounds of interviews, experiencing nonresponse at each of the five rounds. Procedures for making adjustments to compensate for nonresponse bias in the estimation of population parameters have been implemented in the data processing.

Since each annual MEPS sample is a subsample of respondents to the previous year's NHIS, a wide range of survey variables for all the MEPS sampled units (respondents and nonrespondents) are available from the MEPS sampling frame, the NHIS. A set of 19 NHIS variables were identified in 1996, the first year of MEPS, as relevant to response propensity and have been used since then as potential covariates for the first round dwelling unit (DU) level nonresponse adjustment. Each year, a subset of these variables is selected through CHAID (Chi-square Automatic Interaction Detection program) and used for the DU level nonresponse adjustment to the weights for that year's panel. The procedure was documented by Cohen and Machlin (1998). Another 10 variables were identified by Kashihara *et al* (2003) and added to the set in 2004. These 29 variables are listed in Appendix A.

While adjustment variables that are optimal for reducing nonresponse bias should be related to both response propensity and survey outcome variables, the later criterion has not always been taken into consideration when selecting nonresponse adjustment covariates (Kreuter et. al. (2010)) In addition, survey paradata (i.e., information about the survey data collection process) are becoming more widely available and applied to reduce measurement error and biases in survey data. This study incorporates these two considerations as part of a review of the nonresponse adjustment covariates currently used as candidates for the MEPS round 1 DU level nonresponse adjustment. Our analysis is based on the following three-part framework for revising the set of variables currently used: 1) re-assess the association between the current set

¹ This is an exploratory analysis and the options provided in this paper serve as potential alternative strategies to give future consideration to as the set of variables used for the MEPS DU NR adjustments are updated.

of 29 variables and round 1 DU nonresponse, 2) expand the use of NHIS and MEPS paradata for nonresponse adjustment, and 3) evaluate the association between potential nonresponse covariates and selected key MEPS outcome variables. More specifically, we first examine the 29 covariates and eliminate the ones that were not used in either of the two panels (section 2). We then investigate variables from paradata files of the NHIS and MEPS to update the set of potential nonresponse covariates. We selected the variables which were significantly related to response status through regression analysis (section 3). We then put these newly selected variables along with the ones from the current 29 that were retained into a regression model to make a final selection of variables for an updated set of potential covariates (section 4). We then test whether the resulting selected nonresponse covariates are significantly related to key MEPS outcome variables to ensure the adjustment of weights to compensate for nonresponse using those covariates can be shown to effectively reduce bias (section 5). Finally, in section 6 we propose alternative options for revisions to nonresponse adjustment covariates used for the MEPS DU nonresponse adjustment.

Data used in this study were derived from the MEPS panel 12 (P12) round 1 (of year 2007), and panel 13 (P13) round 1 (of year 2008) and their associated NHIS data sets – the 2006 NHIS sample for MEPS panel 12, and 2007 NHIS sample for MEPS panel 13.

2. REVIEW OF THE CURRENT 29 COVARIATES

We examined the 29 potential covariates currently used for DU level nonresponse adjustment. For panels 12 and 13, the following nine were not significant in the selection process by CHAID and therefore were excluded from further consideration in this study:

EMP_REF	Employment status of the reference person
URB_STAT	Urban/Rural residence
TYPE_PSU	Type of primary sampling unit (PSU)
INTVLANG	Interview language
USCITIZN	US citizenship of the reference person
BORNUSA	Born in US – reference person
HOMETYPE	Type of home, e.g., house, apartment etc.
TIMENOPH	Time period without phone – interruption in phone service
HOMEOWN	Homeowner status of the reference person

This leaves 20 of the original 29 currently used variables in this study.

3. SELECTION OF NEW PARADATA COVARIATES

The next step in this investigation involved canvassing the paradata files to identify variables which may be relevant to propensity of response. There were 22 variables initially identified as potentially relevant to propensity of response. However, due to high item nonresponse, lack of a significant chi-square test result, or there being another similar variable already selected, the list was reduced to 10. The 10 new paradata covariates are:

From MEPS Paradata files:

INTNSTY_N2	How strong was the refusal, last
MODE_FRST	Mode of first contact
NHIS_COMP	NHIS completed interview
NUM_CONT	Number of contacts
NUM_INTV	Number of interviewers for the case
NUM_PERS	Number of in-person contacts
NUM_PHONE	Number of phone contacts

From NHIS Paradata files:

RESPOND	How likely to respond to later linked survey
COOPFAM	Assessment of household cooperativeness
ENDPNT	Point in the interview period when the case was finished

Appendix B shows the distributions for these variables as well as chi-square tests of their association with response status.

Variables that were significant by themselves as shown at the previous step were put into logistic regression models to further test their significance. The analysis was done using the stepwise selection method in the SAS logistic procedure. Results of this analysis are given in Appendix C.

As shown in appendix C, the following five variables were significant for both panels:

INTNSTY_N2	How strong was the refusal
MODE_FRST	Mode of first contact
NUM_INTV	Number of interviewers for the case
NUM_PHONE	Number of phone contacts
RESPOND	How likely to respond to later linked survey

The following three variables were significant for one panel:

NUM_CONT	Number of contacts (P12)
NUM_PERS	Number of in-person contacts (P13)
COOPFAM	Assessment of household cooperativeness (P13)

The following two variables were not significant in either panel:

NHIS_COMP	NHIS completed interview
ENDPNT	Point in the interview period when the case was finished

Therefore, the first eight variables are retained as new covariates for nonresponse adjustment.

4. FINAL SELECTION WITH CURRENT AND NEW VARIABLES TOGETHER

For each of the two panels we put the variables that are currently selected and used in the panel along with the newly selected variables into one logistic regression model. The resulting

significant covariates are given in Appendix D. The following nine variables are significant in both panels:

Current variables (among the ones currently used in panel 12 and/or panel 13)

AGE_REF	Age of the reference person
DUSZ_CAT	Number of persons in the DU
HAS_PHONE	Telephone number status in NHIS
REFIONRF	Census region

New paradata variables (among the newly selected in section 3 above)

INTNSTY_N2	How strong was the refusal
MODE_FRST	Mode of first contact
NUM_INTV	Number of interviewers for the case
NUM_PHONE	Number of phone contacts
RESPOND	How likely to respond to later linked survey

The following five are significant in one panel

Current variables (among the ones currently used in panel 12 and/or panel 13)

R_E_REF	race ethnicity of the reference person (P13)
GEOSDIT3	Geographic distribution – MSA size (P13)

New paradata variables (among the newly selected in section 3 above)

NUM_CONT	Number of contacts (P12)
NUM_PERS	Number of in-person contacts (P13)
COOPFAM	Assessment of household cooperativeness (P13)

Altogether these 14 variables constitute the new set of potential covariates for DU level nonresponse adjustment of weights.

5. ASSOCIATION BETWEEN THE SELECTED NONRESPONSE COVARIATES AND OUTCOME VARIABLES

In this section we check to see whether the selected nonresponse covariates are significantly associated with the outcome variables. We selected one outcome variable from each of the five main analytical areas and checked the association between the nonresponse covariates and these outcome variables. The five areas and specific variables include:

Expenditure: TOTEXPy (total health care expenditure for the year yy (yy=07 for panel 12, 08 for panel 13))

Health status: RTHLTH31 (perceived health status)

Access to care: MDUNAB42 (unable to get necessary medical care)

Insurance coverage: INSCOVyy

Utilization: OBTOTVyy (total office based visits)

Coding of these five outcome variables is given in appendix E.

The association of the current 29 variables and outcome variables reflecting the five main analytical areas has been reported in Wun et al (2009). The six current variables (among the 29) selected in section 4 were significantly associated with one or more of the outcome variable as reported in Wun et al (2009). Therefore we need only check the significance of associations between the eight newly selected variables and the five selected.

The results of the association measured by the significance of the chi-square between each pair of variables are provided in Appendix F. The 'x' indicates significance at 0.05 level. All eight new paradata nonresponse covariates selected in section 4 above are significantly associated with one or more of the outcome variables.

6. UPDATING OPTIONS

Based on the results in this study, we can update the set of potential covariates for the DU level adjustment of weights to compensate for nonresponse by replacing the current set of 29 covariates by the set of 14 identified in section 4. Alternatively, the following more conservative options which retain some or all of the current 29 covariates that are not significant in this study can be considered:

1. Keep the 29 currently used variables plus the 8 new variables identified in section 4 above.
2. Replace the 9 currently unused ones (identified in section 2) by the 8 new variables identified in section 4 above.

REFERENCES

Cohen, S.B. and Machlin, S.R. (1998). "Nonresponse Adjustment Strategy in the Household Component of the 1996 Medical Expenditure Panel Survey," *Journal of Economic and Social Measurement*, 25, 15-33.

Kashihara, D., Ezzati-Rice, T.M., Wun, L-M., and Baskin (2003). "An Evaluation of Nonresponse Adjustment Cells for the Household Component of the Medical Expenditure Panel Survey (MEPS)". *2003 Proceedings of the American Statistical Association*, Survey Research Methods Section [CD-ROM], Alexandria, VA: American Statistical Association: 3758-3765.

Kreuter, F., Olsen, K., Wagner, J., Yan, T., Ezzati-Rice, T.M., Casas-Cordero, C., Lemay, M., Peytchev, A., Grove, R.M., and Raghunathan, T.E. "Using proxy measures and other correlates of survey outcomes to adjust for non-response: examples from multiple surveys" *J R Stat Soc Ser A Stat Soc*, 2010 Apr; 173(2): 389-407.

Wun, Lap-Ming, Ezzati-Rice, Trena M. and Kashihara, David (2009)
"Assessment and evaluation of nonresponse error in the Medical Expenditure Panel Survey"
ASA Proceedings of the Joint Statistical Meetings, 2296-2301
American Statistical Association (Alexandria, VA)

Appendix A
Variables currently used in MEPS for nonresponse adjustment

	Variable	Description
1	AGE_REFC	Age category (of the reference person)
2	R_E_REF	Same as R_E_STAT, race ethnicity (of the reference person)
3	ANYASIAN	1 if Any Asian in HH
4	PREDPOV	1 if Any RU w/ Pred Poverty > .3
5	ANYBLACK	1 if no Asian, no Poor, at least one Black person in HH
6	MARRYREF	Marital status of the DU reference person
7	SEX_REF	Gender
8	DUSZ_CAT	Count of the number of persons in a DU
9	ED_DU	Education level of the DU reference person
10	INC_REF	Family income of the DU reference person
11	EMP_REF	Employment status of the DU reference person
12	HAS_FONE	Telephone number status in NHIS
13	HEALTHDU	Same as HEALTHDU, health status in the DU
14	HELP_DU	Does the DU reference person need help with personal care needs
15	REGIONRF	Census region
16	GEODIST3	Geographic distribution – MSA size
17	MSA_STAT	Same as MSA – MSA/Non MSA residence
18	URB_STAT	Same as URB_RRL, Urban/Rural residence
19	TYPE_PSU	Type of PSU
20	HOMETYPE	Type of Home
21	TIMENOPH	Time without a Telephone
22	INTVLANG	Interview language
23	USCITIZN	U.S. Citizen
24	MEDEXPND	Family Medical Expenses Amount
25	HOMEOWN	Homeowner status
26	BORNUSA	Born in U.S.
27	REASONNW	Reason did not Work Last Week
28	HOSPNITE	Number of Nights in the Hospital Last year
29	HLTHCOVR	Health Care Coverage Status

Appendix B - profile of additional variables (NHIS and MEPS paradata)

Variable	Panel 12			Panel 13		
	Overall n=7294	Responding n=5633	Non- responding n=1661	Overall n=9688	Responding n=7575	Non- responding n=2113
	Percent (standard error)			Percent (standard error)		
<u>VARIABLES FROM MEPS PARADATA</u>						
INTNSTY_N2	Chi-square=558.95, p<0.0001			Chi-square=421.64, p<0.0001		
mild (no hostility)	3.85 (0.23)	2.80 (0.22)	7.41 (0.64)	1.29 (0.11)	0.12 (0.04)	5.49 (0.50)
determined (but not hostile)	10.26 (0.36)	3.98 (0.26)	31.55 (1.14)	5.96 (0.24)	0.75 (0.10)	24.61 (0.94)
firm (some hostile)	6.54 (0.29)	1.70 (0.17)	22.94 (1.03)	4.17 (0.20)	0.29 (0.06)	18.08 (0.84)
adamant (hostile)	4.47 (0.24)	0.43 (0.09)	18.18 (0.95)	3.51 (0.19)	0.11 (0.04)	15.71 (0.79)
other (e.g., not refuse)	74.88 (0.51)	91.09 (0.38)	19.93 (0.98)	94.98 (0.36)	98.61 (0.13)	36.11 (1.04)
missing	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.09 (0.03)	0.12 (0.04)	0.00 (0.00)
MODE_FRST	Chi-square=119.33, p<0.0001			Chi-square=167.90, p<0.0001		
mail	3.45 (0.21)	3.21 (0.23)	4.27 (0.50)	0.92 (0.10)	0.75 (0.10)	1.51 (0.27)
in person	69.45 (0.54)	66.29 (0.63)	80.19 (0.98)	69.99 (0.47)	66.65 (0.54)	81.97 (0.84)
telephone	22.96 (0.49)	25.19 (0.58)	15.41 (0.89)	23.45 (0.43)	25.50 (0.50)	16.09 (0.80)
missing	4.13 (0.23)	5.31 (0.30)	0.12 (0.09)	5.64 (0.23)	7.09 (0.29)	0.43 (0.14)
NHIS_COMP	Chi-square=133.49, p<0.0001			Chi-square=89.24, p<0.0001		
complete	80.53 (0.46)	83.83 (0.49)	69.36 (1.13)	75.12 (0.44)	78.43 (0.47)	63.27 (1.05)
partial complete	19.47 (0.46)	16.17 (0.49)	30.64 (1.13)	24.58 (0.44)	21.45 (0.47)	36.73 (1.05)
missing				0.09 (0.03)	0.12 (0.04)	0.00 (0.00)
NUM_CONT	Chi-square=558.40, p<0.0001			Chi-square=473.09, p<0.0001		
1 - 2	23.75 (0.50)	30.46 (0.61)	0.96 (0.24)	22.22 (0.42)	27.72 (0.51)	2.51 (0.34)
3 - 5	24.21 (0.50)	27.55 (0.60)	12.88 (0.82)	31.81 (0.47)	36.46 (0.55)	15.14 (0.78)
6 - 10	21.14 (0.48)	19.37 (0.53)	27.15 (1.09)	23.91 (0.43)	21.73 (0.47)	31.71 (1.01)
11+	30.90 (0.54)	22.62 (0.56)	59.00 (1.21)	21.97 (0.42)	13.97 (0.40)	50.64 (1.09)
missing				0.09 (0.03)	0.12 (0.04)	0.00 (0.00)

Appendix B - profile of additional variables (NHIS and MEPS paradata)

Variable	Panel 12			Panel 13		
	Overall n=7294	Responding n=5633	Non-responding n=1661	Overall n=9688	Responding n=7575	Non-responding n=2113
	Percent (standard error)			Percent (standard error)		
NUM_INTV	Chi-square=417.38, p<0.0001			Chi-square=411.96, p<0.0001		
1	37.74 (0.57)	47.35 (0.67)	5.18 (0.54)	39.01 (0.50)	47.54 (0.57)	8.42 (0.60)
2	29.86 (0.54)	34.40 (0.63)	14.45 (0.86)	34.25 (0.48)	37.57 (0.56)	22.34 (0.91)
3	14.61 (0.41)	10.94 (0.42)	27.09 (1.09)	13.36 (0.35)	9.29 (0.33)	27.92 (0.98)
4 - 5	14.41 (0.41)	6.46 (0.33)	41.36 (1.21)	11.00 (0.32)	4.79 (0.25)	33.27 (1.03)
6+	3.22 (0.21)	0.67 (0.11)	11.86 (0.79)	2.29 (0.15)	0.69 (0.09)	8.05 (0.59)
missing	0.15 (0.05)	0.18 (0.06)	0.06 (0.06)	0.09 (0.03)	0.12 (0.04)	0.00 (0.00)
NUM_PERS	Chi-square=470.14, p<0.0001			Chi-square=307.86, p<0.0001		
0 - 1	30.29 (0.54)	38.52 (0.65)	2.35 (0.37)	17.50 (0.39)	21.45 (0.47)	3.31 (0.39)
2 - 3	21.66 (0.48)	24.94 (0.58)	10.54 (0.75)	32.86 (0.48)	38.06 (0.56)	14.20 (0.76)
4 - 5	14.11 (0.41)	13.40 (0.45)	16.50 (0.91)	17.67 (0.39)	17.65 (0.44)	17.75 (0.83)
6 - 10	18.73 (0.46)	14.45 (0.47)	33.23 (1.16)	19.51 (0.40)	15.37 (0.41)	34.36 (1.03)
11+	15.22 (0.42)	8.68 (0.38)	37.39 (1.19)	12.38 (0.33)	7.35 (0.30)	30.38 (1.00)
missing				0.09 (0.03)	0.12 (0.04)	0.00 (0.00)
NUM_PHONE	Chi-square=104.15, p<0.0001			Chi-square=146.99, p<0.0001		
0	33.88 (0.55)	38.61 (0.65)	17.82 (0.94)	35.95 (0.49)	41.24 (0.57)	16.99 (0.82)
1	21.94 (0.48)	22.81 (0.56)	18.96 (0.96)	23.41 (0.43)	24.61 (0.49)	19.12 (0.86)
2 - 3	20.94 (0.48)	20.06 (0.53)	23.90 (1.05)	20.85 (0.41)	19.63 (0.46)	25.22 (0.94)
4 - 5	9.78 (0.35)	8.61 (0.37)	13.73 (0.84)	8.56 (0.28)	6.77 (0.29)	14.96 (0.78)
6+	13.48 (0.40)	9.91 (0.40)	25.59 (1.07)	11.14 (0.32)	7.63 (0.31)	23.71 (0.93)
missing				0.09 (0.03)	0.12 (0.04)	0.00 (0.00)

Appendix B - profile of additional variables (NHIS and MEPS paradata)

Variable	Panel 12			Panel 13		
	Overall n=7294	Responding n=5633	Non-responding n=1661	Overall n=9688	Responding n=7575	Non-responding n=2113
	Percent (standard error)			Percent (standard error)		
<u>VARIABLES FROM NHIS PARADATA</u>						
RESPOND	Chi-square=70.49, p<0.0001			Chi-square=66.55, p<0.0001		
definitely agree	40.24 (0.57)	43.94 (0.66)	27.69 (1.10)	40.23 (0.50)	43.30 (0.57)	29.20 (0.99)
probably agree	43.05 (0.58)	43.69 (0.66)	40.88 (1.21)	43.31 (0.50)	43.72 (0.57)	41.84 (1.07)
probably refuse	12.31 (0.38)	9.80 (0.40)	20.83 (1.00)	13.32 (0.35)	10.83 (0.36)	22.24 (0.90)
definitely refuse	3.11 (0.20)	1.78 (0.18)	7.65 (0.65)	2.76 (0.17)	1.85 (0.15)	6.01 (0.52)
missing	1.29 (0.13)	0.80 (0.12)	2.95 (0.42)	0.39 (0.06)	0.30 (0.06)	0.71 (0.18)
COOPFAM	Chi-square=62.35, p<0.0001			Chi-square=58.60, p<0.0001		
very good	57.01 (0.58)	61.05 (0.65)	43.29 (1.22)	57.92 (0.50)	61.48 (0.56)	45.15 (1.08)
good	27.24 (0.52)	27.25 (0.59)	27.21 (1.09)	27.10 (0.45)	26.61 (0.51)	28.82 (0.99)
fair	9.72 (0.35)	8.04 (0.36)	15.41 (0.89)	10.34 (0.31)	8.51 (0.32)	16.90 (0.82)
poor	4.74 (0.25)	2.86 (0.22)	11.14 (0.77)	4.25 (0.21)	3.09 (0.20)	8.42 (0.60)
missing	1.29 (0.13)	0.80 (0.12)	2.95 (0.42)	0.39 (0.06)	0.30 (0.06)	0.71 (0.18)
ENDPNT	Chi-square=40.80, p<0.0001			Chi-square=23.13, p<0.0001		
early	35.52 (0.56)	38.27 (0.65)	26.19 (1.08)	32.03 (0.47)	33.50 (0.54)	26.74 (0.96)
middle	31.05 (0.54)	31.07 (0.62)	31.01 (1.13)	31.34 (0.47)	31.93 (0.54)	29.20 (0.99)
late	32.33 (0.55)	30.02 (0.61)	40.16 (1.20)	36.27 (0.49)	34.31 (0.55)	43.30 (1.08)
missing	1.10 (0.12)	0.64 (0.11)	2.65 (0.39)	0.36 (0.06)	0.25 (0.06)	0.76 (0.19)

Appendix C

Results of stepwise logistic regression (new paradata variables only)

Variable name	Panel 12		Panel 13	
	Chi-square	P-value	Chi-square	P-value
<u>MEPS paradata</u>				
INTNSTY_N2	901.90	<0.0001	1124.74	<0.0001
MODE_FRST	38.08	<0.0001	27.49	<0.0001
NHIS_COMP *				
NUM_CONT	95.04	<0.0001		
NUM_INTV	268.52	<0.0001	253.35	<0.0001
NUM_PERS			89.12	<0.0001
NUM_PHONE	21.47	0.0003	129.38	<0.0001
<u>NHIS paradata</u>				
RESPOND	54.24	<0.0001	15.05	0.0046
COOPFAM			21.90	<0.0001
ENDPNT *				

* Not significant in either panel; SAS stepwise method does not provide final statistics of non-significant covariates.

Appendix D

Results of stepwise logistic regression (current and new variables combined)

Variable name	Panel 12		Panel 13	
	Chi-square	P-value	Chi-square	P-value
<u>MEPS paradata</u>				
INTNSTY_N2	858.72	<0.0001	1080.42	<0.0001
MODE_FRST	36.58	<0.0001	24.35	<0.0001
NUM_CONT	92.98	<0.0001		
NUM_INTV	269.55	<0.0001	243.51	<0.0001
NUM_PERS			70.27	<0.0001
NUM_PHONE	21.77	0.0002	135.13	<0.0001
<u>NHIS paradata</u>				
RESPOND	33.93	<0.0001	10.18	0.0375
COOPFAM			16.69	0.0008
<u>Current NR Covariates</u>				
AGE_REFC	24.90	<0.0001	11.60	0.0206
R_E_REF			17.25	0.0006
ANYASIAN*				
PREDPOV*				
ANYBLACK*				
MARRYREF*				
SEX_REF*				
DUSZ_CAT	9.59	0.0479	15.08	0.0045
ED_DU*				
INC_REF				
HAS_FONE	14.90	0.0019	25.93	<0.0001
HEALTHDU*				
HELP_DU*				
REGIONRF	8.18	0.0424	19.45	0.0002
GEODIST3			16.20	0.0003
MSA_STAT*				
MEEXPND*				
REASONNW*				
HOSPNITE*				
HLTHCOVR*				

* Not significant in either panel; SAS stepwise method does not provide final statistics of non-significant covariates.

Appendix E

Re-coded categories of the outcome variables:

TOTEXPy_{yy} (total health care expenditure of the year yy (yy=07 for panel 12, 08 for panel 13))

Categories of this variable are created by SAS PROC RANK. Ten categories are created with cutoff points are the percentiles of 10%, 20%, ..., 90%. Each category has approximate same number of observations.

RTHLTH31 (perceived health status)

- 0 other (inapplicable, refused, don't know, or not ascertained)
- 1 excellent
- 2 very good
- 3 good
- 4 fair
- 5 poor

MDUNAB42 (unable to get necessary medical care)

- 0 other (inapplicable, refused, don't know, or not ascertained)
- 1 yes
- 2 no

INCOV_{yy}

- 1 any private
- 2 public only
- 3 uninsured

OBTOTV_{yy} (total # office based visits)

- 1 0 or 1
- 2 2
- 3 3
- 4 4 – 5
- 5 6 – 8
- 6 9 – 15
- 7 More than 15

Appendix F

Table of association between the eight paradata variables and the five outcome variables

	Outcome areas/variables									
	Expenditures		Health Status		Access to care		Insurance coverage		Use (Office-based visit)	
<u>NR_cov</u> ↓	totexp07	totexp08	rthlth31 (p12)	rthlth31 (p13)	mdunab42 (p12)	mdunab42 (P13)	inscov07	inscov08	obtotv07	obtotv08
RESPOND	X	X	X	X	X	X	X			
COOPFAM	X	X	X	X	X	X	X	X		
INTNSTY_N2			X	X	X	X				
MODE_FRST	X	X	X	X	X	X	X	X		X
NUM_CONT	X	X	X	X	X	X	X	X	X	
NUM_INTV		X	X	X	X	X		X		
NUM_PERS	X	X	X	X	X	X	X	X	X	X
NUM_PHONE		X	X	X	X	X	X	X		

X = Chi-square test significant at $\alpha=0.05$ level