STATISTICS IN MEDICINE
Statist. Med. 2007; **26**:1875–1884
Published online 5 January 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2809

On modelling response propensity for dwelling unit (DU) level non-response adjustment in the Medical Expenditure Panel Survey (MEPS)[§]

Lap-Ming Wun[1,*,†,‡], Trena M. Ezzati-Rice[1], Nuria Diaz-Tena[2] and Janet Greenblatt[1]

[1]*Agency for Healthcare Research and Quality, Rockville, MD, U.S.A.*
[2]*Mathematica Policy Research, Inc., U.S.A.*

SUMMARY

Non-response is a common problem in household sample surveys. The Medical Expenditure Panel Survey (MEPS), sponsored by the Agency for Healthcare Research and Quality (AHRQ), is a complex national probability sample survey. The survey is designed to produce annual national and regional estimates of health-care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian non-institutionalized population. The MEPS sample is a sub-sample of respondents to the prior year's National Health Interview Survey (NHIS) conducted by the National Center for Health Statistics (NCHS). The MEPS, like most sample surveys, experiences unit, or total, non-response despite intensive efforts to maximize response rates. This paper summarizes research on comparing alternative approaches for modelling response propensity to compensate for dwelling unit (DU), i.e. household level non-response in the MEPS.

Non-response in sample surveys is usually compensated for by some form of weighting adjustment to reduce the bias in survey estimates. To compensate for potential bias in survey estimates in the MEPS, two separate non-response adjustments are carried out. The first is an adjustment for DU level non-response at the round one interview to account for non-response among those households subsampled from NHIS for the MEPS. The second non-response adjustment is a person level adjustment to compensate for attrition across the five rounds of data collection. This paper deals only with the DU level non-response adjustment. Currently, the categorical search tree algorithm method, the chi-squared automatic interaction detector (CHAID), is used to model the response probability at the DU level and to create the non-response adjustment cells. In this study, we investigate an alternative approach, i.e. logistic regression to model the response probability. Main effects models and models with interaction terms are both evaluated. We further examine inclusion of the base weights as a covariate in the logistic models. We compare variability of weights of the two alternative response propensity approaches as well as direct use of propensity scores.

[*]Correspondence to: Lap-Ming Wun, Center for Financing, Access and Cost Trends (CFACT), Agency for Healthcare Research and Quality (AHRQ), 540 Gaither Road, Room 5240, Rockville, MD 20850, U.S.A.

[†]E-mail: lwun@ahrq.gov

[‡]Mathematical Statistician.

[§]This article is a U.S. Government work and is in the public domain in the U.S.A.

Received 15 October 2006
Accepted 16 November 2006

Published in 2007 by John Wiley & Sons, Ltd.

The logistic regression approaches produce results similar to CHAID; however, using propensity scores from logistic models with interaction terms to form five classification groups for weight adjustment appears to perform best in terms of limiting variability and bias. Published in 2007 by John Wiley & Sons, Ltd.

KEY WORDS: propensity score; weighting; logistic regression; relative total variation

1. INTRODUCTION: MEPS SURVEY DESIGN AND ESTIMATION STRATEGY

The Medical Expenditure Panel Survey (MEPS) is a complex national probability sample survey sponsored by the Agency for Healthcare Research and Quality (AHRQ). MEPS is designed to provide nationally representative estimates of health-care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian non-institutionalized population. The MEPS consists of three inter-related surveys with the household component (HC) as the core survey. The sample for the MEPS-HC is drawn from respondents to the National Health Interview Survey (NHIS), conducted by the National Center for Health Statistics (NCHS). The MEPS-HC uses an overlapping panel design in which data are collected through a series of five rounds of interviews over a two and one-half year period. Detailed information on the MEPS sample design has been previously published [1, 2].

The MEPS-HC, like most sample surveys, experiences unit, or total, non-response despite intensive efforts to maximize response rates. Two separate non-response adjustments are performed as part of the process for development of analytic weights in MEPS. The first is an adjustment for dwelling unit (DU) non-response at round 1 to account for non-response among those households subsampled from NHIS for the MEPS. The 1996–2000 MEPS DU response rates ranged from 80 to 83 per cent (among the NHIS households fielded for MEPS). The second adjustment for non-response is at the person level to account for survey attrition across the various rounds of data collection. The other adjustment carried out in the development of weights in MEPS is poststratification of the non-response adjusted weights to known population counts, i.e. estimates of the population from the Current Population Survey (CPS). The focus of this paper is on the DU non-response adjustment. The investigation reported here was conducted with the round 1 non-response of the 2000 MEPS.

2. NON-RESPONSE WEIGHTING ADJUSTMENT

Survey non-response is usually compensated for by some form of weighting adjustment to reduce the potential bias in survey estimates. The use of classifying or auxiliary variables, i.e. covariates, to form non-response adjustment cells is a commonly used method for non-response adjustment. It has been shown by Cochran [3] that it is effective in removing non-response bias in observational studies. Rosenbaum and Rubin [4] have indicated that as the number of covariates increases, the number of classes grows exponentially and suggested using predicted response probabilities (i.e. propensity scores) from a logistic regression model based on a set of covariates to form the weighting classes or cells. Another adjustment method is the use of the inverse of the respondent's predicted propensity score as an adjustment factor [5]. In this paper, we call this latter method the 'direct use' of propensity scores. A propensity score of response in surveys is essentially the conditional probability that a person or household responds given the covariates. More elaboration of the propensity score and its application in non-response adjustments can be found in Little [6] and Little and Rubin [7] among others. A previous comparison of the use of covariates *versus* the use of response propensities to form classes for non-response adjustment for a complex sample survey, the third National Health and Nutrition Examination Survey (NHANES III), was reported by Ezzati-Rice and Khare [8].

The current method implemented by the MEPS contractor, Westat, to compensate for DU level non-response in the MEPS is based on chi-square automatic interaction detection (CHAID) models to divide the DUs into non-response adjustment cells [9]. CHAID is one version of the automatic interaction detector (AID) developed for categorical variables. The underlying theory for the CHAID has been discussed by Kass [10]. In brief, CHAID is an exploratory method to examine the relationship between a dependent variable (e.g. non-response) and a series of predictor variables and their interactions. The CHAID algorithm creates adjustment cells by splitting a data set progressively *via* a classification tree structure where the most important predictor variables are chosen that maximize a chi-square criterion. The most significant predictors define the first split or the first branching of the tree. Progressive splits from the initial variables result in smaller and smaller branches. The result at the end of the tree building process is a series of groups that are different from one another on the dependent variable. A Bonferroni type adjustment is used to correct for the number of different ways a single predictor variable can be split [11]. The CHAID methodology has been used for adjusting for non-response in other surveys [12].

In this paper, as contrast to the CHAID method, we investigate four separate logistic models for predicting the probability of response and three alternative ways of using the probabilities calculated from each of the four models to adjust weights to compensate for non-response. The primary research objective of this study is to determine whether methods of using response propensity scores calculated from logistic regression models have any important advantages as compared to the current CHAID method. The comparison is carried out to address two main questions: (1) Whether a simple, straightforward, main effects only logistic model is sufficient for non-response adjustment in MEPS. If so, the steps of developing, testing, and merging branches of the classification tree using CHAID or model identification using more sophisticate regression models can be skipped and (2) Whether more sophisticated models with interaction terms are needed. If more sophisticated models are needed, CHAID and model identification cannot be avoided. But, the regression model-based method would still be simpler to implement than the current method, as well as being more efficient as suggested by Rosenbaum and Rubin [4]. As part of the modelling efforts, we also consider the issue of weighted and unweighted logistic models. The 'weighted' approach follows what Little and Vartivarian [13] suggest, namely the inclusion of the base weights as a covariate in the logistic model rather than running a traditional weighted regression. The weights used in our logistic model are the MEPS base weights. Specifically, the base weight in the MEPS is the reciprocal of an intermediate weight from the NHIS reflecting the disproportionate sampling of minorities in NHIS with a ratio adjustment to NHIS population estimates to account for NHIS non-response and undercoverage.

The logistic regression models use the same set of covariates as currently used in CHAID. For many surveys, little information is available about the non-respondents. However, the MEPS has a unique advantage with a relatively large amount of information available for the non-respondents resulting from the MEPS use of a subsample of respondents to the previous year's NHIS. For this study, we also report the results of poststratification of the non-response adjusted weights. Poststratification is the step that follows the non-response adjustment in the development of weights in the MEPS. The DU level non-response adjusted weights are poststratified at the family level to match the CPS control totals.

3. ALTERNATIVE METHODS OF NON-RESPONSE ADJUSTMENT

In this study, we compare the various methods of DU non-response adjustment following round 1 of the 2000 MEPS. The DU non-response at the round 1 of the 2000 MEPS was 19 per cent. Currently, Westat uses a tree diagram generated by the computer software package CHAID to form non-response adjustment cells based on response propensity. There are 17 covariates used in the procedure. Cells are collapsed, if necessary, to ensure that the number of respondents in a cell are no less than 20 (Göksel *et al.* [14] 'MEPS Panel V Round 1—DU level weights' unpublished internal memo of Westat, WGTS # 469.01, P5R1 #1.01). The 17 variables listed below are input to CHAID as potential predictors of response propensity to construct subclasses for the DU non-response adjustment in the 2000 MEPS-HC. These classifying variables were identified based on analysis of 1996 MEPS-HC data and are based on information collected in the NHIS [15].

1. Age of the DU reference person.
2. Race/ethnicity of the DU reference person.
3. Marital status of the DU reference person.
4. Gender of the DU reference person.
5. Number of persons in the DU.
6. Education of the DU reference person.
7. Family income of the reference person.
8. Employment status of the DU reference person.
9. Phone number refused in NHIS.
10. Major work status—working or reason for not working.
11. DU level health status.
12. DU reference person needs help with daily activities.
13. Census region.
14. Metropolitan Statistical Area (MSA) size.
15. MSA/non-MSA status of the DU.
16. Urban/rural status of the DU.
17. Type of primary sampling unit (PSU)—self representing *versus* non-self representing.

An alternative to the current CHAID propensity non-response adjustment method is to develop a logistic regression model to predict response status using a set of covariates. A propensity score of response is essentially the conditional probability of response given the covariates. For this study, it was calculated through the following steps:

1. Run a logistic regression with response/non-response indicator as the dependent variable on a set of covariates based on the 17 predictor variables as currently used in CHAID.
2. Convert the estimated logit value obtained from the logistic model established in step 1 into the predicted probability of response, i.e. the propensity score, through the following equation:

$$\mathrm{PROB} = \mathrm{EXP}(\mathrm{LOGIT})/(1 + \mathrm{EXP}(\mathrm{LOGIT}))$$

DOI: 10.1002/sim

With a propensity score calculated for each sample unit, the propensity scores from the logistic regression were used in two different ways in this study:

1. *Direct*: The estimated propensity score of each respondent was used directly as the adjustment factor, i.e. each individual respondent's base weight was multiplied by the inverse of their propensity score.
2. *Grouping of scores to form adjustment cells*: Using the propensity scores, the sample was grouped into classification cells. In this study, we report the results from groupings of 5 and 100. The selection of 5 groups was based on the optimality established by Cochran [3] and extended to propensity scores in observational studies by Rosenbaum and Rubin [4]. These studies showed that 5 classes were often sufficient to remove 90 per cent of the bias due to the covariates. The inclusion of 100 groups was designed to assess the effect of a much finer classification of the propensity scores while keeping the number of respondents in a cell at no fewer than 20 to match the criterion used by Westat in the current CHAID method.

4. LOGISTIC REGRESSION MODELS

For our research questions related to: (1) whether a simple, straightforward, main effects only logistic model is sufficient or (2) whether more sophisticated models with interaction terms are needed, along with the issue of weighted and unweighted modelling for the logistic models, we develop and evaluate the following four logistic regression models:

Model 1: use of the 17 covariates currently used in the CHAID method.
Model 2: use of the 17 covariates currently used in CHAID plus the base weight (total of 18 covariates).
Model 3: use of the 17 covariates currently used in CHAID and interaction terms to identify 'best' set of potential predictors.
Model 4: use of the 17 covariates currently used in CHAID plus the base weight and interaction terms to identify the 'best' set of potential predictors.

Models 1 and 2 simply use all 17 covariates in the regression. Model 1 is the unweighted one and model 2 is the weighted version. In models 3 and 4, since some of the 17 auxiliary variables associated with response status might be correlated with each other and thus one of the two might be adequate for use in the adjustment and since there might also be combinations of items that could predict response status, we carried out the following steps to identify a best set of response status variables:

Step 1: Use CHAID to model the non-response pattern, identify main effects and interactions.
Step 2: Check correlations between the main effects to eliminate collinearity.
Step 3: Run logistic regression to identify significant terms.

From the initial 17 covariates included in model 3, six main effect variables and one interaction term were selected in the final reduced model. The selected terms were:

- Education of the reference person.
- Income of the reference person.
- Whether there is a working phone inside the house and the phone number is provided to the interviewer.
- Employment status of the reference person.

Published in 2007 by John Wiley & Sons, Ltd.

DOI: 10.1002/sim

- Census region.
- Martial status of the reference person.
- Interaction of income and region.

From the 17 initial covariates along with the base weight included in model 4, six main effects (one being the base weight), and one 2-way interaction, and one 3-way interaction were selected in the final reduced model. The selected terms were:

- Education of the reference person.
- Income of the reference person.
- Whether there is a working phone inside the house and the phone number is provided to the interviewer.
- Employment status of the reference person.
- Census region.
- Base weight.

and interactions of:

- Income and region.
- Phone, income, and region.

With the covariates for each of the models selected, each model was put through the steps of calculating the propensity score. The propensity scores were then used to adjust the weights to compensate for non-response as described in the previous section.

5. POSTSTRATIFICATION

At the last step, the DU non-response adjusted weights based on each of the non-response adjustment methods (cell weighting and direct use of propensity score) under the selected models were poststratified at the family level to totals obtained from the March 2000 CPS. The poststratification was done within classes formed by family type, race/ethnicity, region, MSA status, age, and family size.

6. EVALUATION CRITERIA AND RESULTS

A major objective of adjusting weights to compensate for non-response is to have the adjusted weights of the respondents representing the original population total; that is, the sum of the adjusted weights for respondents equals the sum of the pre-adjustment weights (the base weights in this study), i.e. before any non-response adjustment. Therefore, if the set of adjusted weights from the same model/method has a sum that equals the sum of the base weights of all sample units, then they are considered unbiased, and the mean of this set is designated as the true mean of adjusted weights. The methods of adjusting weights in classification cells or groups of propensity scores adjust the weights of respondents in the cell by the ratio of the sum of the base weights of all units in the cell to the sum of the base weights of respondents in the cell. The sum of the resulting adjusted weights equals to the sum of all the base weights in the same cell. Therefore, the mean of adjusted weights from the method of grouping under any of the models is unbiased. The mean of adjusted weights from the method of using the propensity score directly usually does not equal the true mean, i.e. it is usually biased, which is the nature of a model based approach; there is always some discrepancy between the modelled value and the true value in regression analysis. With the true mean and unbiasedness defined here, we use the following measure, which we shall call relative total variation (RelTV), to compare results from each model/method:

$$\text{RelTV} = \text{CV} + \text{relative bias} = (\text{standard deviation}/\text{mean}) + (|\text{mean} - \text{true mean}|/\text{mean})$$

RelTV is the sum of the coefficient of variation (CV) and the relative bias. The relative bias is the ratio of the absolute value of the difference between the mean of the given set of weights and the true mean to the mean. Hence, for a set of weights with unbiased mean, the RelTV is their CV.

Summary statistics of the distribution of the adjusted weights are given in Tables I and II. Table I shows the RelTV and the mean of the non-response adjusted weights based on each of the models with each of the three alternative methods of using the propensity scores generated from logistic regressions. For example, the cell in the row of Model 1 and the column labelled direct is the statistics for the set of weights adjusted using propensity scores as calculated from logistic model 1 and with the propensity scores used directly (i.e. use of the inverse of the propensity score as the adjustment factor). Table II shows the results after the poststratification of the non-response adjusted weights of the model/methods that were selected after comparing results in Table I.

In Table I, the mean of 25 053 is the 'true' mean as we have defined above. All sets of adjusted weights from either of the 5-group or the 100-group methods under each of the four models have this unbiased value. The means from the direct method are more or less different from this true value. Among them, the ones from models 1 and 2 are much further from the true value than those from models 3 and 4. Also the RelTVs from the direct method under models 1 and 2 are much higher than those from any other model/method. Therefore, we decided not to investigate these two models any further, and only carry models 3 and 4 to the next stage of poststratification. It can also be noted that the mean from the current CHAID method is also slightly different from the unbiased number of 25 053. This is because after the non-response adjustment, the non-response

Table I. Relative total variation (in percentage) and mean (in parentheses) of DU level non-response adjusted weights by model and propensity score method.

	Propensity score methods			
Model	CHAID	Direct	5 groups	100 groups
Current	43.68 per cent (24 738)	NA	NA	NA
Model 1 (17 cov.)	NA	64.19 per cent (20 058)	42.95 per cent (25 053)	44.30 per cent (25 053)
Model 2 (17 cov. + base weight)	NA	64.19 per cent (20 058)	44.21 per cent (25 053)	45.77 per cent (25 053)
Model 3 (new + interaction)	NA	42.86 per cent (24 810)	41.71 per cent (25 053)	42.92 per cent (25 053)
Model 4 (new + interaction + base weight)	NA	44.56 per cent (24 997)	43.19 per cent (25 053)	44.55 per cent (25 053)

Note: Current/CHAID = Method currently used in MEPS using cell classification modeled by CHAID. Columns 3–5: Direct = method of using propensity score directly. 5 groups = method of using propensity scores to classify units into 5 groups. 100 groups = method of using propensity scores to classify units into 100 groups.

Table II. Relative total variation of DU level non-response adjusted and family level post-stratified weights by selected model and propensity score method.

	Propensity score methods			
Model	CHAID	Direct	5 groups	100 groups
Current	52.06 per cent	NA	NA	NA
Model 3 (new + interaction)	NA	51.23 per cent	50.88 per cent	51.86 per cent
Model 4 (new + interaction + base weight)	NA	52.91 per cent	51.64 per cent	53.34 per cent

Note: The mean is 25 534 for all cases after post-stratification. Current/CHAID = Method currently used in MEPS using cell classification modeled by CHAID. Columns 3–5: Direct = method of using propensity score directly. 5 groups = method of using propensity scores to classify units into 5 groups. 100 groups = method of using propensity scores to classify units into 100 groups.

adjusted weights of the DUs which are found to be ineligible for MEPS are set to zero. For the purpose of this evaluation, the true unbiased number is needed; thus, we kept the non-zero weights from all other methods for those DUs.

In Table II, after poststratification, the mean, as shown in the footnote, is of the same value of 25 534 under each model/method. The RelTV of the distribution of the weights from model 3 using propensity scores to form 5 groups is the smallest among all sets of non-response adjusted and poststratified weights. Therefore, we interpret it as the best in the sense that it is unbiased and has the smallest relative total variation among all sets of non-response adjusted and poststratified weights from the various model/methods evaluated.

7. DISCUSSION

Weights are commonly developed for use in analysis of data collected in sample surveys. A survey's analytical weights typically account for any disproportionate probabilities of selection, unit non-response, as well as an adjustment to make the weighted sample distributions agree with known population distributions for selected demographic variables. This paper focused on the first step of research into alternative adjustment methods to compensate for DU non-response in the MEPS. Since the MEPS sample is linked to the NHIS, a large number of potential variables are available from the NHIS for MEPS respondents and non-respondents. This study used a previously identified set of 17 auxiliary variables and compared two primary methods for non-response adjustments: CHAID (a categorical tree algorithm approach) and logistic regression. The propensity scores obtained from the logistic regression models were used in three alternative ways: direct and groupings of scores into 5 and 100 groups. The logistic models included main effects models and models with interaction terms. The effect of the inclusion of the base weight as a covariate in the logistic models was also evaluated.

This study only examined the variability of the weights for the alternative methods and models as one measure of the effect of the adjustments on the precision of the survey estimates. The means from the direct use of propensity scores for model 1 (based on current set of 17 variables)

DOI: 10.1002/sim

and model 2 (current set of 17 variables and base weight) were much smaller than the 'true' mean. This indicates a serious bias in the adjusted weights using these two models. Hence, the goal of using the simplest adjustment method, i.e. propensity scores directly from a main effects model, was not promising. Under the enhanced models 3 and 4, i.e. models with interactions and the base weight included, the method of 5 groups (of the propensity scores) appeared to be optimal, as Cochran suggested. The method in which the propensity scores were aggregated into 100 groups had the higher RelTV (it is the same as the CV for a distribution with unbiased mean) than the RelTV from the 5 groups in each model. This indicates refinement of propensity scores groups beyond the optimal level (i.e. 5 groups) may have undesirable results. This is not surprising since larger number of groups results in smaller number of sample units in each group. The proportion of respondents in some of the groups may be much smaller than in any of the groups under the 5-group classification; this would result in larger adjustment factors (multipliers) and lead to increased variance. The RelTVs of the non-response and poststratified weights from model 3 (reduced model with interactions) were slightly smaller then those from model 4 (reduced model with interactions and inclusion of the base weight). This may be due to the fact that model 4 has more terms in the model (8 terms) than model 3 (7 terms). More independent terms in a regression model often lead to more variation. In any case, this result indicates that the base weight (DUPSWT) does not make much difference in modelling the response propensity in this current data set analysis. Overall, the models comprised of a reduced set of variables with interactions, i.e. models 3 and 4, and using the resulting propensity scores to form 5 groups for adjusting weights were the better performer in this study based on examination of the variability of the weights.

This research was only the first phase of an ongoing research effort focused on weighting adjustments in the MEPS. Future research will examine a new 'best' set of auxiliary variables based on the large pool of potential NHIS data available for both MEPS respondents and non-respondents. Rizzo *et al.* [12] report that the choice of auxiliary variables may be the most important step of the weighting adjustment for non-response. In addition, we will assess alternative adjustment methods based on the identified enhanced set of covariates and in particular the effectiveness of alternative variables and adjustment methodologies in reducing non-response bias will be evaluated for a range of MEPS estimates. Finally, the research will be carried forward to the person level survey attrition in the MEPS.

REFERENCES

1. Cohen SB. Sample design of the medical expenditure panel survey household component. *Agency for Health Care Policy and Research*, *MEPS Methodology Report*, *No. 2*, AHCPR Publication No. 97-0027, Rockville, MD, 1997.
2. Cohen SB. Sample design of the 1997 Medical Expenditure Panel Survey household component. *Agency for Healthcare Research and Quality*, *MEPS Methodology Report*, *No. 11*, AHRQ Publication No. 01-0001. Rockville, MD, 2000.
3. Cochran WG. Removing bias in observational studies. *Biometrics* 1968; **24**:295–313.
4. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on propensity score. *Journal of the American Statistical Association* 1984; **79**:516–524.
5. Kalton G, Flores-Cervantes I. Weighting methods. *Journal of Official Statistics* 2003; **19**(2):81–97.
6. Little RJA. Survey nonresponse adjustments for estimates of means. *International Statistical Review* 1986; **54**:139–157.
7. Little RJA, Rubin DB. *Statistical Analysis with Missing Data* (2nd edn). Wiley: New York, 2002.
8. Ezzati-Rice TM, Khare M. Modeling of response propensity in the third National Health and Nutrition Examination Survey. *Proceedings of Survey Research Methods Section of the American Statistical Association* 1994; 955–959.

9. Cohen SB, DiGaetano R, Goksel H. Estimation procedures in the 1996 Medical Expenditure Panel Survey household component. *Agency for Health Care Policy and Research*, *MEPS Methodology Report No. 5*, AHCPR Publication No. 99-0027, 1999.
10. Kass GV. An exploratory technique for investigating large quantities of categorical data. *Applied Statistics* 1980; **29**:119–127.
11. The Measurement Group. Available at http://www.themeasurementgroup.com/Definitions/chaid.htm (accessed April 26, 2006).
12. Rizzo L, Kalton G, Brick JM. A comparison of some weighting adjustment methods for panel nonresponse. *Survey Methodology* 1996; **22**:43–53.
13. Little RJ, Vartivarian S. On weighting the rates in non-response weights. *Statistics in Medicine* 2003; **22**: 1589–1599.
14. Göksel HA, Alvarez-Rojas L, Hao H. MEPS Panel V Round 1—DU level weights. Unpublished internal memo of Westat, *WGTS # 469.01*, *P5R1 # 1.01*, 10 December 2001.
15. Cohen SB, Machlin SR. Nonresponse adjustment strategy in the household component of the 1996 Medical Expenditure Panel Survey. *Journal of Economic and Social Measurement* 1998; **25**:15–33.

Published in 2007 by John Wiley & Sons, Ltd.

DOI: 10.1002/sim