Investigation of Random Effects Models for Time Varying Variables with Missing Values in the Medical Expenditure Panel Survey (MEPS)

Robert M. Baskin[1], Sourish Saha[2], and John Fleishman [1]
Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850[1]
University of Florida[2]

Keywords: random effects; PWIGLS; missing at random; drop out missingness

1. Introduction

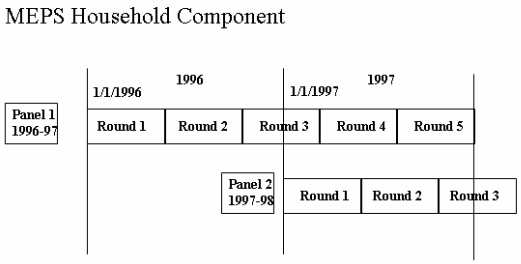
The Medical Expenditure Panel Survey (MEPS) is a complex national probability sample survey sponsored by the Agency for Healthcare Research and Quality (AHRQ). MEPS is designed to provide nationally representative estimates of health care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian noninstitutionalized population. MEPS consists of a family of three interrelated surveys with the Household Component (HC) as the core survey. MEPS-HC is an overlapping panel survey and eligible respondents remain in the survey for five consecutive rounds of data collection. The fact that MEPS-HC is collected in ongoing panels provides the advantage of performing longitudinal analysis on data but also introduces the challenges of addressing missing data issues in a longitudinal dataset.

The MEPS-HC, like most sample surveys, experiences item nonresponse despite efforts to collect complete information. The amount of nonresponse does vary by variable. Healthcare expenditures is considered one of the primary analysis variables in MEPS-HC but many other variables including self reported health status are collected. There is a substantial amount of item nonresponse on expenditures in MEPS-HC but there is usually minimal nonresponse within a round on self reported health status. But dropout missingness, while not necessarily large in percent of observations, can have substantial impact on the results of a longitudinal analysis. To compensate for missing expenditure data, a weighted sequential hotdeck imputation approach is used to reduce the potential bias in estimating expenditures. However, self reported health status is not imputed in MEPS-HC. The purpose of this study is to investigate methods for accounting for missingness, especially dropout missingness, in longitudinal analysis of time varying variables in MEPS-HC such as in self reported health status.

2. Background: MEPS Sample

The sample of households for the MEPS-HC is a subsample of households that responded to the prior year's National Health Interview Survey, conducted by the National Center for Health Statistics. 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. The standard representation for the time periods of the round of data collection for the first two MEPS panels is displayed in Figure 1.

Figure 1.



Analytic weights, accounting for survey nonresponse, are calculated for MEPS-HC and the details on the weights as well as detailed information on the MEPS-HC sample design can be found in (Cohen, 1997; Cohen, 2000).

Medical expenditures is one of the primary analysis variables collected in the MEPS-HC. It has a high percent of missingness. To compensate for the missing data and to improve the accuracy of the survey estimates, data on expenses for household respondents are also collected from a sample of their health care providers in the Medical Provider Component of MEPS. However, frequently expense data are not available from either survey. If expenditure values are not available from either survey component, then it is imputed.

Self reported health status is typically not a primary analytic variable but is often used as auxiliary information in model building. It is not imputed, so if there is missingness on this variable for some time periods, then the missingness has to be dealt with by the end user.

3. Drop Out Missingness

Drop out missingness is a major concern in longitudinal datasets. In many cases dropouts are different from responders in ways that nonresponse adjustment or full case analysis cannot account for. Also note that for good reasons imputation is rarely used for drop outs, thus imputation is not a viable solution to drop out missingness. For an overview of nonresponse and drop out missingness in particular, see Little and Rubin (2002).

Self reported health status is a time varying variable in MEPS that is a good variable for a study of techniques associated with drop out missingness. It is not imputed and does not have much item missingness. It also does not have the distributional problems associated with expenditures such as skewness and semicontinuity.

Self reported health status is a five point Likert scale variable with 1 corresponding to Excellent health and 5 corresponding to Poor health. Self reported health status has some occasional missingness not associated with drop outs but for longitudinal analysis self reported health status does show patterns of both drop outs and drop ins. This is represented in Figure 2.

Figure 2.
Missingess Patterns

After initial collection in round 1 only about 1% of the subjects drop out in either round 2 or in round 4. In round 3 it appears that there is about 6% drop out cases but this is an exaggeration of the actual dropouts and is probably due to an artifact of processing. There are also some nonresponders who can fail to provide a response for a given period but return with a response in a later period. Even though dropouts represent a very small percentage of the total cases, they cannot be ignored because the subjects who are dropping out can be very different from the general population. For example, the mean expenditures in round 1 for the subjects who dropped out after round 1 is about eight times as large as the mean expenditures in round 1 for the rest of the subjects. The implications of this are that the dropouts are seriously different and not accounting for them could bias any analytical results. This idea is shown for expenditures and self reported health status in Table 1.

Table 1: All Persons

		Total Sample	Drop Outs	Not Drop Outs
Total Health Care Ex-pend-itures	Sample Size	21574	128	21446
	Weighted Size (millions)	269	2	267
	Mean	$2,037	$15,300	$1,941
	SE Mean	$67.29	$3,202	$64.56
Self Reported Health Status	Sample Size	21221	79	21142
	Mean	2.11	3.3	2.11
	SE Mean	0.01	0.18	0.01

source of data: 1996 MEPS

Table 2 provides the mean of self reported health status by round for the overall group and broken down by whether the subject dropped out at the following round. It is further indication of the difference of drop outs from the responders.

Table 2

Health Status for Panel 1	Overall Mean of Health Status	Not Missing at Following Round	Missing at Following Round	P-Value for Difference
Round 1	2.1133	2.1075	3.2854	0
Round 2	2.1164	2.1056	3.4626	0
Round 3	2.1195	2.124	2.0825	0.1808
Round 4	2.1349	2.1313	3.0054	0

Source of data: 1996 MEPS

4. Research Strategy

There are numerous references, such as Skinner, Holt, and Smith (1989), which point out the dangers of ignoring the complex sampling design in analyzing data from a complex survey, so for the purposes of this paper it is assumed that this point needs no further discussion. All of the methods investigated here take the complex design into account in some form. It is also strongly argued in Skinner, Holt and Smith (1989) that the traditional design based approach, which they refer to as p-weighting, while possibly producing valid variance estimates in the sense that the design is accounted for, rarely produces valid

inferences of interest for models.

Chambers and Skinner (2003), which is the follow up volume to Skinner, Holt, and Smith (1989), review the latest methods for attempting to make valid inferences with complex survey data. Skinner and Holmes (2003) specifically address building random effects models on longitudinal data from complex surveys. In Little and Rubin (2002), random effects models are shown, under relatively reasonable assumptions, to allow the user to assume missing data is missing at random as opposed to missing completely at random. This relationship between random effects models and assuming missing at random is strongly dependent on the likelihood approach. To transfer this property of being able to assume missing at random to the complex design setting would require fundamentally a likelihood approach that incorporated some form of complex design validity.

This is precisely the approach in Skinner and Holmes (2003) and this approach will be investigated here for the purpose of building random effects models for five rounds of self reported health status in MEPS-HC.

Skinner and Holmes (2003) also discuss an approach that more closely resembles a design based approach but requires the existence of 'wave' weights which are not currently produced for MEPS. For this reason the approach was not pursued.

An alternative model for comparison purposes is described in chapter ten of Skinner, Holt, and Smith (1989). This alternative approach for building models is to use a model based variance estimate but to incorporate design information such as strata and cluster variables as covariates in the model. This approach can be criticized on the grounds that it can ignore information in the sampling weights if the design is informative. However, from an analytic point of view it does provide valid inferences.

5. Random Effects Models for Longitudinal Survey Data

As previously mentioned, Skinner and Holmes (2003) address the problem of building random effects models based on complex survey data. This method was developed in a sequence of papers by Pfeffermann and Lavange (1989), Pfeffermann, et al, (1998), and Skinner and Holmes (2003). This sequence of papers takes an evolutionary path starting from Generalized Least Squares (GLS) to probability weighted GLS to Iterative Generalized Least Squares (IGLS) to probability weighted Iterative Generalized Least Squares (PWIGLS). This path is evolutionary in the sense of the modern quality movement in statistics and references for each step can be found in Skinner and Holmes (2003).

A sketch of the ideas behind PWIGLS is as follows. If we want to build a random effects model on our data, then the model will have regression parameters, denoted by $\boldsymbol{\beta}$, and variance components denoted by $\boldsymbol{\theta}$. If we start by ignoring any design weights and build a model using IGLS then we make some initial guess at $\boldsymbol{\theta}$, say $\boldsymbol{\theta}_0$. Using the value of $\boldsymbol{\theta}_0$, we produce an estimate of $\boldsymbol{\beta}$, say $\boldsymbol{\beta}_1$ using generalized least squares. Now, treating $\boldsymbol{\beta}_1$ as a constant go back and produce a new estimate of $\boldsymbol{\theta}$, say $\boldsymbol{\theta}_1$. This process is iterated, updating each set of parameters consecutively until convergence is achieved.

At this point the method looks promising because it is a likelihood based approach but so far it doesn't account for the sample weights. The next step is to incorporate the sample weights. Pfeffermann, et al, (1998) modified the IGLS method so that sample weights are included in the following way. First the weights are normalized in a clever way described in Pfeffermann, et al, (1998). Then the normalized weights are included in the IGLS random effects model as a covariate but not as design weights (p-weights). This produces point estimates of the $\boldsymbol{\beta}$ that are simultaneously model unbiased and design consistent. Unfortunately, if the design is not equally weighted (self-weighting) this modification produces estimates of variance that are neither model consistent nor design consistent, so one further modification is required. At this point in the second step of the IGLS approach a design consistent estimator of variance needs to be employed. This has the unfortunate effect that standard software for IGLS has to be stopped at each iteration and a new variance estimate inserted into the process or that the iteration has to done by hand. Nevertheless, this method does provide a way of fitting random effects models for complex survey data that is likelihood based but still design consistent.

There is one further modification in Skinner and Holmes (2003) which accounts for the longitudinal aspect of data. The adjustment is made to account for serially correlated data in the longitudinal model. Altogether this produces models with the desired theoretical properties but the aspect of actually fitting the model is quite tedious.

One further technical aspect of the method that needs clarification is how to determine the original estimate of variance. Pfeffermann, et al, (1998) suggest starting with a design consistent estimator of the variance components $\boldsymbol{\theta}_0$ conditional on a design consistent estimate of $\boldsymbol{\beta}$ from a standard design based model.

6. Model Construction

In this study, models were fit using PWIGLS with self reported health status as the dependent variable and some

demographic variables such as age, education level, MSA, and race as predictor variables. Age and Education level were treated as ordinal whereas Race and MSA were treated as categorical. Because of the size of the dataset, models were only fit for females in Census Region 1 at the first round. The repeated observation over time was treated as the random effect.

There are several limitations to this model. The dependent variable is ordinal but treated as approximately normal. Models were fit only for females in Census Region 1 and only for Panel 1 which is 1996 – 1997 data. The information for segment sampling is unavailable and this is considered a major loss of information for many variables of interest. It might be of interest to treat person as random in the model but this is totally impractical in terms of the capability of the software to invert a design matrix with person as random.

In spite of the limitations, the following five models, increasing in computational complexity, were fit for self reported health status. Model A is a model based estimate with fixed demographic regressors only. Model B is the same as Model A except with strata and PSU variables added as regressors. Model C is Model B with a random effect added for repeated observation over time. Model D is 'almost' PWIGLS in that the second step of the iterative process did not replace the estimate of variance with a design consistent estimate. This model is worth considering because it saves considerable difficulty in the fitting of the true PWIGLS model. The final model, Model E, is full PWIGLS. These models are outlined in Table 2.

(insert table 3.)

7. Model Fit

The parameters for the models are given in Table 4 at the end. The variables used in all models are Age, Education Level, Race and MSA. Age and Education were treated as ordinal and centered at their mean. Race and MSA were treated as categorical with two levels for MSA and five levels for Race. The design information was entered as PSUxxxxxx with strata and PSU level encoded and this was treated as categorical. The normalized weight variables were called wiggle1 and wiggle2 and are standardized to have mean 0 and variance 1. The variable called round is the random effect and gives the round of collection. The variable round has a small parameter which is never significant but it does switch signs between the fixed effects and mixed effects models. The other common parameters appear to be of similar magnitude across the models although if the models are theoretically different it is not clear that this is actual information.

The residuals for each model were investigated. The random effects models are standardized with smaller error so they are not on the scale as the fixed effects models. Nevertheless, something of interest did occur in the residual table. The full PWIGLS model has a median shift in residuals compared to the other two random effects models. See the Residuals in Table 5 at the end. This raises a theoretical question of how to judge the residuals from a PWIGLS model. It is not clear if the residuals from the PWIGLS model should be looked at as weighted residuals or if there is some underlying problem here. Perhaps this is a sign of a programming error.

8. Discussion

Five different models were fit to the data and the model fit seemed to improve with each increase in model complexity. However, the residuals seem to indicate a median shift for the PWIGLS model. This could possibly be because the residuals for a PWIGLS model should be weighted by the normalized sampling weights used in the PWIGLS model. This is a theoretical question for which the authors currently do not have an answer.

However, there is another consideration. The PWIGLS model is a random effects model that allows the end user to assume the data is missing at random. The models that are not random effects models force the end user to assume the data is missing completely at random. Under this set of circumstances, should the residuals or parameters from the two types of models even be compared? We have seen evidence that the drop outs for self reported health status are not missing completely at random so this would seem to be a spurious comparison at this point.

Finally, the difficulty in fitting the PWIGLS model cannot be overemphasized. It is tedious enough so that some type of error in creating the estimates was likely to occur. This type of error is extremely dangerous and an automated method of producing the estimates needs to be devised.

9. Conclusion

This study attempted to determine if it was possible to fit random effects models for longitudinal data in MEPS-HC. The current state of theory and technology make this seem possible but the reality of the fitting is that it is extremely tedious to go through this process.

The next step in this project for future research could be to attempt to automate the iterative fitting. At this point the only statistically valid alternative to PWIGLS is fitting a Bayesian finite population model.

References

Chambers, R. and Skinner, C.J. (2003) Analysis of Survey Data, New York: Wiley.

Cohen S.B. (1997). "Sample Design of the Medical

Expenditure Panel Survey Household Component". Agency for Health Care Policy and Research, MEPS Methodology Report, No. 2, Rockville, MD., AHCPR Pub. No. 97-0027.

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

Little, R.J.A. and Rubin, D.B. (2002) Statistical Analysis With Missing Data (2nd Ed.). New York: Wiley.

Pfeffermann, D. and Lavange, L. (1989), "Regression Models for Stratified Multi-Stage Cluster Samples", in Analysis of Complex Surveys (Skinner, C. J., Holt, D., and Smith T.M.F. eds.), chapter 12, New York: Wiley.

Pfeffermann, D., Skinner, C.J., Holmes, D.J., Goldstein, H. and Rasbash, J. (1998), "Weighting for Unequal Selection Probabilities in Multi-Level Models", Journal of the Royal Statistical Society, Series B, 60, 23-40.

Skinner, C.J. and Holmes, D.J. (2003), "Random Effects Models for Longitudinal Survey Data", in Analysis of Survey Data (Chambers, R. and Skinner, C.J. eds.), chapter 14, New York: Wiley.

Skinner, C. J., Holt, D., and Smith T.M.F. (1989) Analysis of Complex Surveys, New York: Wiley.

Table 3.

Model Designation	Fixed Demo-graphic Re-gressors	Design Re-gressors	Random Effects for Repeated Measures	Design Consistent PWIGLS	p-value
A	x				
B	x	x			<0.0001
C	x	x	X		<0.0001
D	x		X	step A Only	<0.0001 #
E	x		X	Step A&B	<0.0001 #

This test may not be valid. Skinner and Holmes use a Rao-Scott adjustment for it.

Table 4A: Model A

	Estimate	Std. Error	t value
(Intercept)	1.833393	0.244493	7.499
round	-0.00169	0.007332	-0.23
MSA961	0.112189	0.032839	3.416
RACEX2	1.308938	0.389025	3.365
RACEX3	0.439375	0.249484	1.761
RACEX4	0.516635	0.242837	2.127
RACEX5	0.366419	0.241497	1.517
I(EDUCYR96 - 9.5)	-0.03957	0.002587	-15.3
I(AGE96X - 46)	0.020025	0.000588	34.068

Table 4B: Model B

	Estimate	Std. Error	t value
(Intercept)	1.677527	0.294604	5.694
round	-0.00214	0.007298	-0.294
MSA961	0.273533	0.165274	1.655
RACEX2	1.098236	0.390079	2.815
RACEX3	0.41581	0.249592	1.666
RACEX4	0.511354	0.243276	2.102
RACEX5	0.3262	0.241903	1.348
I(EDUCYR96 - 9.5)	-0.03881	0.002595	-14.956
I(AGE96X - 46)	0.019708	0.000592	33.27

PSU96022	0.011155	0.029563	0.377
PSU96023	-0.09941	0.045367	-2.191
PSU96021004	0.2521	0.051322	4.912
PSU96021005	-0.08731	0.05006	-1.744
PSU96021006	0.125365	0.073931	1.696
PSU96021007	0.162417	0.063902	2.542
PSU96021008	0.073637	0.08177	0.901
PSU96021009	0.155055	0.176031	0.881
PSU96021010	0.154832	0.180849	0.856
PSU96021011	0.151159	0.187535	0.806
PSU96021012	0.459419	0.12579	3.652
PSU96021013	0.030798	0.103654	0.297
PSU96021014	0.107623	0.232529	0.463
PSU96021047	0.194718	0.430107	0.453
PSU96021056	-0.00875	0.215784	-0.041
PSU96021094	0.415497	0.430417	0.965
PSU96023004	-0.06572	0.049791	-1.32
PSU96023005	0.081726	0.059783	1.367
PSU96023006	-0.18749	0.082248	-2.28
PSU96023007	0.255271	0.071117	3.589
PSU96023008	0.105525	0.111112	0.95
PSU96023009	0.219729	0.180337	1.218
PSU96023010	-0.05678	0.191332	-0.297
PSU96023011	0.313098	0.194645	1.609
PSU96023012	0.082203	0.104644	0.786
PSU96023013	0.303996	0.125853	2.415
PSU96023014	0.429479	0.185177	2.319
PSU96023047	0.497175	0.434594	1.144
PSU96023048	-0.58983	0.430764	-1.369
PSU96023087	0.940925	0.429746	2.189

Table 4C: Model C			
	Value	Std.Error	t-value
(Intercept)	1.788965	0.341009	5.2461
Round	0.00276	0.007381	0.373933
MSA961	0.137003	0.190692	0.718452
RACEX2	0.852745	0.485176	1.7576
RACEX3	0.509178	0.290981	1.74987
RACEX4	0.55009	0.282802	1.94514
RACEX5	0.346847	0.281008	1.2343
I(EDUCYR96 - 9.5)	-0.03955	0.003277	-12.068
I(AGE96X - 46)	0.020289	0.000746	27.1987
PSU96022	0.012157	0.037463	0.324509
PSU96023	-0.14403	0.057926	-2.4864
PSU96021004	0.196165	0.065522	2.99387
PSU96021005	-0.0602	0.062234	-0.96737
PSU96021006	0.196378	0.094119	2.08649
PSU96021007	0.245541	0.081125	3.0267
PSU96021008	0.107289	0.09979	1.07514
PSU96021009	-0.00891	0.205649	-0.04333
PSU96021010	0.089262	0.212619	0.419823
PSU96021011	-0.00415	0.217127	-0.01913
PSU96021012	0.451086	0.162557	2.77493
PSU96021013	0.150385	0.133913	1.12301

PSU96021014	-0.1423	0.284608	-0.5
PSU96021047	0.204373	0.556017	0.367567
PSU96021056	-0.09416	0.278924	-0.33758
PSU96021094	0.768986	0.556428	1.38201
PSU96023004	-0.05424	0.062989	-0.86117
PSU96023005	0.023689	0.076082	0.31136
PSU96023006	-0.12869	0.101947	-10.2623
PSU96023007	0.184427	0.089149	2.06875
PSU96023008	0.042666	0.14193	0.300611
PSU96023009	0.05142	0.211921	0.242636
PSU96023010	-0.21606	0.225956	-0.95623
PSU96023011	0.177975	0.230782	0.771184
PSU96023012	0.079052	0.133804	0.590802
PSU96023013	0.347068	0.162628	2.13412
PSU96023014	0.318143	0.216782	1.46757
PSU96023047	0.174498	0.561387	0.310835
PSU96023048	-0.35411	0.556853	-0.63591
PSU96023087	1.188468	0.555575	2.13917

Table 4D: Model D

	Value	Std.Error	t-value
(Intercept)	1.7008	0.29271	5.81062
round	0.01104	0.01003	1.100026
MSA961	0.12653	0.04566	2.771088
RACEX2	1.08079	0.48396	2.233321
RACEX3	0.57528	0.29062	1.979465
RACEX4	0.58886	0.28183	2.08941
RACEX5	0.42523	0.28024	1.517394
I(EDUCYR96 - 9.5)	-0.0397	0.00333	-11.9086
I(AGE96X - 46)	0.02036	0.00074	27.52686
wiggle1	-2.26704	5.05085	-0.44884
wiggle2	66.31474	31.80067	2.085326

Table 4E: Model E: available from the contact author upon request

Table 5.

		Residuals:				
		Min	1Q	Median	3Q	Max
Fixed	A	-2.5244	-0.8199	-0.0249	0.6921	3.38647
only	B	-2.4791	-0.8038	-0.0308	0.64631	3.40101
		Standardized Within-Group Residuals:				
Random	C	-2.5568	-0.6643	-0.0726	0.54783	3.66516
Effects	D	-2.416	-0.6538	-0.0665	0.54303	3.84292
	E	-2.7566	-0.6125	-0.1653	0.63268	3.46439