Using Repeated Measures Techniques To Analyze Cluster-correlated Survey Responses

G. Gordon Brown, Celia R. Eicheldinger, and James R. Chromy
RTI International, Research Triangle Park, NC 27709

Abstract

The researchers for the CAHPS® of Medicare Fee-For-Service and Disenrollee/Assessment surveys have used a repeated measures approach in SUDAAN to create individual level composite scores on a number of constructs (e.g. the CAHPS composite "Getting needed Health Care" combines four survey questions that ask a beneficiary about experiences with getting care needed). This approach, applicable to any survey analysis with cluster-correlated responses, estimates models that assume each question in the composite is a repeated measure. Advantages include models that adjust for covariates, account for the survey design, and perform hypothesis testing while preserving the CAHPS composite design. Traditionally the CAHPS composites are analyzed using the CAHPS Macro; however, the CAHPS Macro does not possess the ability to correctly conduct the hypothesis tests we required. Our goals are to 1) disseminate this useful technique to researchers using the CAHPS methodology and 2) generalize the technique to other cluster-correlated survey responses.

Keywords: Repeated Measure, CAHPS®, SUDAAN

1. Introduction

Analyzing data that come from a complex survey presents a number of challenges to the researcher. These challenges can be amplified significantly when the data come from a sample survey that has a longitudinal or repeated measures component. Software that is designed for sample surveys, such as SUDAAN (Research Triangle Institute, 2004), may not be specifically designed to address the longitudinal aspect of a repeated measures design. Software that does have the appropriate procedures (survey or repeated measures) may not have the capability of correctly adjusting the variances for the design of the data. The researcher is often left pondering the best course of action. In some of these situations, an optimal or near optimal solution can be found, which brings us to the focus of this paper: (1) produce variance estimates that are adjusted for the sample design and a longitudinal or repeated measures component of a survey and (2) provide a method to conduct complex hypothesis tests for data with these features.

In a repeated measures or longitudinal data setting, there is often a high degree of correlation between the observations from a given experimental unit, which will affect the variance estimates. Failing to appropriately account for this correlation and to account for the sample design can lead to erroneous conclusions as both the repeated measures aspect of the study and the sample design affect the variances needed for hypothesis testing. Correctly accounting for all types of potential correlation is crucial for any form of statistical analysis. We hope to give the reader an appreciation of the methods needed to account for longitudinal data coming from a complex sample design when performing data analysis.

Any set of survey questions or experimental measurements can be treated as repeated measures if they are, in some fashion, measuring the same outcome or if a combination of the questions or measurements is attempting to measure a single outcome. Often in a survey setting, such as the CAHPS®[1] Medicare Fee-For-Service (MFFS) survey (Centers for Medicare & Medicaid Services [CMS], 2000), this single outcome is attempting to measure an experience based on a serious of three to four related survey questions. Although we have applied this technique to analyzing the CAHPS MFFS data, the models and concepts presented in this paper can easily be applied to any survey or other study with a repeated measures component.

The motivation behind our method, which we call the Repeated Measures CAHPS (RM-CAHPS), is the need to perform trend analyses and conduct complex hypothesis tests when comparisons of more than two group-level means are desired. The specific analysis required on CAHPS MFFS survey data consists of analyzing combinations of questions commonly referred to as composites. The CAHPS Survey Users Network Web site contains a listing of all the composites (Agency for Healthcare Research and Quality [AHRQ], 2004).

[1] Formerly know as the Consumer Assessment of Health Plans Study.

The standard CAHPS analysis method treats these composites as a grouping of questions that together measure the same trait; similar to the repeated measure of a given trait. The ultimate goal is to obtain estimates, called composite scores, that are means for Medicare plans or means for the levels of the analysis variable of interest based on a function of the means of the questions that comprise the composite. These composite scores are then used to conduct various hypothesis tests.

Through out this paper we discuss the advantages and assumptions of using the repeated measures approach to modeling the CAHPS composites, which we call RM-CAHPS. Although we use the CAHPS composites as an example, this technique can be applied to a variety of settings. For example in a panel study where the same people respond to a set of survey questions in multiple years, using the repeated measures approach described in this paper will allow the research to correctly account for the design of the survey as well as account for the within person correlation.

2. Accounting for Sample Design

One of the key aspects of analyzing survey data is accounting for the weighting and sample design when estimating the variance. Most large surveys have analysis weights plus stratification and clustering in the sample design. All of these elements of a complex survey impact the variance and need to be accounted for when conducting data analysis.

Weighting has an impact on both the point estimates and the variance estimates. For this paper, we treat the analysis weights as survey weights and not counts. The typical purpose of survey weights is to indicate the number of persons a given observation represents in the population. They do not indicate the number of observations in the sample, which implies that the sum of the weights represent the number of persons in the population, not the number of persons in the sample. Standard statistical software packages, not designed for survey analysis, usually treat the weights as counts. Consequently, the point estimates are correct but the variances can be grossly underestimated.

Clustered sampling is very common in sample surveys and fairly common in experiments or observational data. In most sample surveys, the correlation within clusters is positive (overdispersion). Failing to account for this correlation will result in a variance estimate that is biased low. Low variance estimates will result in confidence intervals that are too narrow and hypothesis tests that have an inflated Type I error rate (too liberal). Infrequently, the intracluster correlation will be negative (underdispersion) and will produce the opposite effects of overdispersion.

Stratification is often implemented as a method of improving the efficiency of drawing the actual sample. However, effect usages of stratification can lead to a decrease in the estimate of the sampling variance. If homogeneity of responses exists within strata, the estimates of variance can be less than if stratification did not occur. Typically, stratification has less impact on variance estimates than weighting or clustering.

To obtain unbiased variance estimates that are adjusted for sample design features, it is important to use a software package specifically designed for surveys. For example, SUDAAN, SAS (PROC SURVEYREG and PROC SURVEYLOG), STATA, and WESTVAR are software packages that can correctly handle survey data. We used SUDAAN exclusively for our analyses.

3. Correlation within Repeated Measures and CAHPS Composites

In most situations, the observations from repeated measures data within a person or experimental unit are positively correlated. Ignoring this correlation will likely result in variance estimates that are biased low and, consequently, confidence intervals based on these variance estimates that will be too narrow. Similarly, the potential for rejecting a true null hypothesis in favor of an unsupported alternative hypothesis (Type I error) increases. For obvious reasons, this will lead to poor inference and erroneous conclusions with regard to hypothesis tests. For the CAHPS composites, each respondent in a CAHPS survey has the potential to answer each question in a given composite. It is logical to assume that their responses to the questions that make up a given CAHPS composite will be positively correlated.

4. Missing Data

Often, some responses within an experimental unit comprising the repeated measure or questions within a person comprising a CAHPS composite will be missing. For the following reasons, missing observations do not pose any difficulty to the proposed Repeated Measures Method (RM-CAHPS specifically for the CAHPS composites).

Repeated measures designs have a correlation matrix for observations within the experimental unit or cluster where the structure depends on the ordering of the responses. In most analyses, this correlation matrix is considered a nuisance parameter; it needs to be determined to obtain meaningful inference from the analysis, but the actual values of the matrix are not important to the researcher. The use of Generalized Estimating Equations (GEE) (Binder 1983, Zeger and Liang 1986) or other robust variance estimation algorithms alleviates this problem; estimation of the exact correlation structure is unnecessary when using GEE to adjust variance estimates for the sample design. Each cluster is assumed to have a unique correlation structure that does not need to be estimated in order to adjust the estimates of variance. The end result is that the persons or experimental units with missing values do not need to be removed from the study or adjusted in any way.

Another concern is the use of analysis weights with the missing observations. The analysis weights are almost always created for a given person; they are not created for given response items or questions from the survey. However, the repeated measures models that we propose assign weight by response items that contain nonmissing data, not by person. To analyze repeated measures data using the method purposed, a separate record must be created for each response item for each person. For example, if there are four repeated measures per person, there will be four records on the data set for every person. (Please see the section on setting-up the repeated measures for CAHPS MFFS data below.) Each record from a given person will use the analysis weight that is assigned to that person. As a result, a given person will have more impact on the final results when they provide nonmissing responses to more items within the repeated measure than one who provides nonmissing responses to fewer items within the measure. The missing item within the measure is not used in the modeling process; however, other valid items from the same person are used in the modeling process. This allows the use of partial data from a given individual in all of the modeling situations. The only time that a person's data would not be used in the modeling process is if that individual had missing responses to all items that comprise the repeated measure or composite. We must note that missing data for the independent variables in the model will result in the record being deleted during the modeling process.

When using GEE, the degrees of freedom are not affected by missing data. The default degrees of freedom used in our models is the number of design strata minus the number of primary sampling units (PSU). All records for a given person must be in the same PSU to take advantage of the robust variance estimation methods. Therefore, the number of degrees of freedom will not depend on the number of items in a PSU or on missing values. This is an important property of our modeling scheme, as the creation of multiple records per person could artificially inflate the degrees of freedom and once again, provide test results that are too liberal.

5. Modeling Applications

The method that we describe can be employed for a variety of model structures. We will present examples using SUDAAN's linear regression procedure (PROC REGRESS). However, the method is not limited to linear regression models. Other types of models are available in SUDAAN: logistic (PROC LOGISTIC), multinomial (PROC MULTILOG), count (PROC LOGLINK), and survival (PROC SURVIVAL). For all of the point and variance estimates reported, we used the GEE option and a with-replacement design. However, the methods presented here would work using other designs (e.g., without replacement), and other robust variance estimation methods like Replicate Weight Jackknife and Balanced Repeated Replication.

6. Test Statistics

Another issue that we considered is which test statistic to use for multiple degrees-of-freedom hypothesis tests. The standard Wald chi-square test statistic is often too liberal for survey data for multiple degrees-of-freedom tests. For our analyses purposes, we chose the adjusted Wald F (Fellegi 1980), a more appropriate test statistic for survey data. Satterthwaite's adjusted chi-square (Rao and Scott 1981) and the Wald F (RTI 2004, pp.175–177) are other test statistics that are recommended for surveys. These test statistics are available for all procedures in SUDAAN.

7. Methods

The RM-CAHPS begins by fitting a general regression model to the data of interest and was first proposed by Chromy and McLeod (2000). The subgroup variable of interest, covariates, and an indicator for item (question number) are placed in the model. If desired, interactions between independent variables (subgroup and/or covariates) are also included in the model. Once the researcher has the desired terms in the model, to replicate the CAHPS Macro, an addition set of terms must be added. These additional terms are the interaction between all independent variables and the item variable. This interaction has to be included because the CAHPS Macro models each question separately. Note that if the researcher does not care to replicate the CAHPS Macro, the additional interactions are not needed. If the "repeated measures" is truly a repeated measure and not a set of survey questions measuring a similar trait, then the indicator for item is not necessary either. For example, in a panel study if the same question was answered by the same person once every year, then an indicator for year would not be required.

A general form of the model is easily derived. The notation being used is consistent with notation for a complex survey. For the model below, we assume that there is one level of stratification and one primary sampling unit.

$$\mathbf{y}_{ijk} = g(\alpha + subgroup_{ijk} + qa\_num_{ijk} + subgroup * qa\_num_{ijk} + \boldsymbol{\beta}' qa\_num * \mathbf{x}_{ijk}) + \boldsymbol{\varepsilon}_{ijk}$$

$i = 1, 2, \ldots I$ represents the strata variable
$j = 1, 2, \ldots J_i$ represents the PSU within strata
$k = 1,2,\ldots n_{ij}$ represents person or experimental unit within PSU

$$Var(\boldsymbol{\varepsilon}_{ijk}) = \sigma^2 \mathbf{I}_n$$

The response vector $\mathbf{y}_{ijk}$ is composed of all of the responses to the questions that form the composites or repeated measure; in the case of the ratings on single response variable, the response is a scalar. The term $\mathbf{x}_{ijk}$ is a vector of all covariates that are used in the model, and $\boldsymbol{\beta}$ is the corresponding vector of regression coefficients. The vector $\boldsymbol{\varepsilon}_{ijk}$ comprises the residuals. Note that there is a residual for every element of $\mathbf{y}_{ijk}$. The dimensions of the intracluster covariance matrix $Var(\boldsymbol{\varepsilon}_{ijk})$ depend on the number of items for a composite. For a ratings question, $Var(\boldsymbol{\varepsilon}_{ijk})$ is the scalar σ^2. The functional form of $g(*)$ depends on the type of regression. From linear regression, $g(*)$ will be the identity link function, for logistic regression it is the logit link function, and for multinomial models it will be the generalized or cumulative logit link function (Agresti, 1990).

The above variance assumption, $Var(\boldsymbol{\varepsilon}_{ijk}) = \sigma^2 \mathbf{I}_n$, is used for parameter estimation only. To account for the intracluster correlation, variance is estimated using generalized estimating equations (GEE) (Binder, 1983; Zeger and Liang 1986). When using GEE, the $Var(\boldsymbol{\varepsilon}_{ijk})$ matrix is unstructured. It is not necessary to actually estimate the correlation structure to obtain an estimate of variance because the parameters are estimated under the assumption of independence and the estimate of variance is robust to this assumption. This relieves the researcher of the responsibility for determining the exact form of the intracluster correlation matrix.

Predicted margins (direct standardization) are used to produce an estimate of the composite scores from our model (Korn and Graubard, 1999). Assume that the variable subgroup has R levels. The predicted margins are calculated for $r = 1, 2, \ldots R$. For a given level of *subgroup, r,* the formula for the predicted margin is

$$\mathrm{E}(Y \mid X, \beta, subgroup = r) = \sum_{ijk} g(\alpha + subgroup^* + qa\_num_{ijk} + subgroup^* * qa\_num_{ijk} + \boldsymbol{\beta}' qa\_num * \mathbf{x}_{ijk})$$

In this situation, the variable subgroup is set to a given level *r*. The *subgroup** indicates that all observations are set to this value for *subgroup*; all other covariates remain unchanged. Typically, the predicted margins are calculated for every level of the subgroup variable. Hypothesis tests can then be conducted on the margins. For the unequal item weights, the predicted marginals produce the point estimates desired. For the equal weights, contrasts of the predicted marginals are needed to produce the desired estimates.

Modification of the repeated measures model for the CAHPS MFFS data is relatively straightforward. For the example presented, we are interested in trends over years. Our subgroup variable will be *year*. The question numbers will be the items that compose a given CAHPS composite. For the CAHPS MFFS data, the individuals that comprise the sample are the primary sampling units. As a result, the index j is dropped from the model. The repeated measures model is

$$\mathbf{y}_{ik} = g(\alpha + year_{ik} + item_{ik} + year * item_{ik} + \boldsymbol{\beta}' item * \mathbf{x}_{ik}) + \boldsymbol{\varepsilon}_{ik} \quad \begin{matrix} i = 1, 2, ..., 1104 \\ k = 1, 2, ..., n_{ij} \end{matrix},$$

where i is an index for the year**geounit* interaction, which is the stratum variable for our models, and k is an index for person within a *year*geounit* stratum. We are primarily interested in estimating trends across years. This will be done using the predicted margins and calculating the appropriate contrast.

8. Setting Up the Repeated Measures for CAHPS MFFS Data

Here we will detail the steps taken to get the CAHPS MFFS data set-up for the repeated measures analysis. We will also explain how the SUDAAN code should be set-up. First it is important to note that each survey respondent has the potential to provide a response for each question comprising a given composite. In Table 1 the lay out for the file is shown. Notice respondent ID 1 has responses to 4 survey questions. In this mock-up example, there are 4 questions comprising the repeated measure or composite in the CAHPS example. Also notice that for each survey question in the composite there is a record in the data set, thus the number of records in the dataset, for a composite with 4 questions, will be 4 times the number of respondents.

Table 1. Example File Layout for Repeated Measures

Respondent ID	Question # for Composite	Response to Question	Subgroup Variable
1	1	3	Year 1
1	2	5	Year 1
1	3	4	Year 1
1	4	2	Year 1
2	1	2	Year 1
2	2	3	Year 1
2	3	3	Year 1
2	4	5	Year 1
1	1	3	Year 2
1	2	5	Year 2
1	3	4	Year 2
1	4	2	Year 2
2	1	1	Year 2
2	2	5	Year 2
2	3	4	Year 2
2	4	5	Year 2
1	1	3	Year 3
1	2	5	Year 3
1	3	3	Year 3
1	4	2	Year 3
2	1	2	Year 3
2	2	1	Year 3
2	3	3	Year 3
2	4	4	Year 3

Once the data is set-up, we are ready to run the analysis in SUDAAN. Setting up the repeated measures aspect in SUDAAN is relatively simple. The primary sampling unit (PSU) identifies the repeated measure. In our example the respondent ID (called "id" in the code below) indicates the survey respondents and thus the repeated measure. In SUDAAN the PSU is generally the second variable in the nest statement following the stratification variable. If you have more than one stratification variable then the PSU variable can be the third, fourth, etc variable in the nest statement. If this is the case the code for identifying the PSU variable is "psulev = position of PSU variable."

In the example code, below, our stratification variable is the cross of year and geounit (yrgeo) and "id" is our PSU. We also include an indicator of the survey question (qanum) in the model as well as all the interactions with qanum. The researcher does not have to include these extra terms and interactions if they are not trying to mimic the CAHPS macro. And in cases of true repeated measures identifying the repeated measure with an indicator variable is not necessary.

```
proc regress  data = tcare_1 design = wr;
nest yrgeo id;
weight postwght;
class qanum year/nofreq;
model response = qanum   year qanum*year
female under65  qanum*(female under65 ) ;
pred_eff  qanum=(0.25  0.25  0.25
0.25)*year=(1 0 0 0)/NAME="Year =
1";
```

```
pred_eff   qanum=(0.25   0.25   0.25
0.25)*year=(0  1  0  0)/NAME="Year  =
2";
pred_eff   qanum=(0.25   0.25   0.25
0.25)*year=(0  0  1  0)/NAME="Year  =
3";
pred_eff   qanum=(0.25   0.25   0.25
0.25)*year=(0  0  0  1)/NAME="Year  =
4";
predmarg year;
```

Our subgroup variable is "year" and we use the "pred_eff" and "predmarg" statements to obtain the compsites scores by year. The "pred_eff" statements are used to calculate the composite scores with equal weighting of the items since we have four questions in the composite each equation is weighted by 0.25. The researcher can chose to weight the questions differently, perhaps using the factor loadings from a principle components analysis to determine how each question should be weighted. The "predmarg" outputs the predicted marginals which are used for the unequal weighting of item scores. In the regression setting this is the same as the least squares means.

9. Results

The results, found in Table 2, compare the CAHPS composite scores and standard errors for the Getting Needed Care Composite using the RM-CAHPS, a regression model based on a simple random sample (SRS) without any repeated measures, and the CAHPS Macro. These two tables also illustrate the difference between the two CAHPS Macro settings we have chosen to investigate: unequal weighting and equal weighting of the items comprising the composite. In both tables, the predicted marginals generated from linear regression models using SUDAAN are found in the column titled "RM Marginals." (The marginals are the same regardless of survey design, so we present them only once). As already noted, the predicted marginals are very similar to the CAHPS Macro scores. The standardization of both independent and dependent variables performed in the CAHPS Macro explains the differences.

The last three columns demonstrate the effect each method has on the standard errors. When the sample design and the clustering from the items within a composite are ignored, the standard errors are smaller, as shown by the standard errors from the SRS regression model ("RM SE-Naïve") and the standard errors from the CAHPS Macro ("CAHPS Macro SE"). The difference in the standard errors between the CAHPS Marco and the RM-CAHPS range from -0.0012 to -0.0023; a 10% to 19% decrease in the standard error. This decrease in the standard error will result in an increase in the Type I error rate.

Table 2. Comparison Estimates with Unequal Weighted Items

Survey Year	RM Marginals	CAHPS Macro Scores	RM SE-Naive	RM SE-Design	CAHPS Macro SE
Unequal Weighted Items					
Year = 1	2.8183	2.8156	0.0099	0.0119	0.01068
Year = 2	2.8670	2.8633	0.0066	0.0084	0.00680
Year = 3	2.8031	2.8020	0.0108	0.0135	0.01130
Year = 4	2.8305	2.8300	0.0097	0.0118	0.00981
Equal Weighted Items					
Year = 1	2.8253	2.8222	0.0092	0.0113	0.01015
Year = 2	2.8686	2.8644	0.0067	0.0084	0.00679
Year = 3	2.7940	2.7926	0.0131	0.0155	0.01324
Year = 4	2.8256	2.8247	0.0115	0.0135	0.01119

10. Conclusions

The flexibility of the modeling method makes the repeated measures approach applicable to a variety of different situations a researcher may encounter. For example, the models we presented can be modified by replacing *year* with any independent variable of interest. Additionally, all other hypothesis tests available in a regular regression setting are available to the researcher. As expected, these additional tests are based on the full model and account for covariates, weights, and the sample design. The repeated measures approach gives the researcher a great deal of flexibility when modeling clustered data often found in a survey setting.

Two additional features of the GEE method are (1) a complicated correlation structure does not need to be specified for the repeated measures models being fit and (2) missing data do not create problems. This is not the case for other statistical packages, such as SAS's PROC MIXED, where the user is required to specify the correlation structure. Also, during our preliminary investigative work, we discovered that PROC MIXED was not capable of handling large numbers of missing values within the repeated measure (the respondent, in our case). As a result, we were never able to obtain a fitted model using PROC MIXED.

Acknowledgements

Performed under contract 500-95-0061/T.O.#7. We would like to thank Edward S. Sekscenski, Larry Campbell, Lisa Carpenter, Jeffrey Laufenberg, and Shulamit Bernard.

References

Agency for Healthcare Research and Quality (AHRQ). 2004. "Topics Covered by the Core Items in the Health Plan Survey" [accessed on November 2, 2005]. Available at: http://www.cahps-sun.org/Products/Healthplan/HP-CoreItems.asp

Agresti, A. 1990. Categorical Data Analysis. New York: Wiley.

Binder, D. A. 1983 "On the Variances of Asymptotically Normal Estimators from Complex Surveys." *International Statistical Review*, 51, 279–92.

Chromy, J. R. and L. D. McLeod 2001. Considerations for the Development of G-CAHPS Composite Scoring Algorithms. Development and Testing of the Group-Level Consumer Assessment of Health Plans Study Instrument: Results from the National G-CAHPS Field Test. G.-C. R. Team. Research Triangle Park, NC, RTI International**:** A110-A129.

Centers for Medicare & Medicaid Services (CMS). 2000. " Implementation of CAHPS® Fee-for-Service Survey—Final Report for Year 1" [accessed on November 3, 2005]. Available at: http://www.cms.hhs.gov/researchers/projects/consumers/ffs_overview1.pdf

Fellegi, I. P. 1980. "Approximate Tests of Independence and Goodness of Fit Based on Stratified Multistage Samples." *American Statistical Association, Proceedings of the Section on Survey Research Methods,* pp. 495–500. Baltimore MD: American Statistical Association.

Korn, E. L., and B. I. Graubard. 1999. Analysis of Health Surveys. New York: Wiley.

Rao, J. N. K., and A. J. Scott. 1981. "The Analysis of Categorical Data from Complex Surveys: Chi-Squared Tests for Goodness of Fit and Independence in Two-Way Tables." *Journal of the American Statistical Association*, 76 (374): 221–30.

Research Triangle Institute. 2004. SUDAAN Language Manual (release 9.0).Research Triangle Park, NC: RTI International. [accessed on November 3, 2005] Available at: http://www.rti.org/sudaan/

Zeger, S. and K. Liang (1986). "Longitudinal data analysis for discrete and continuous outcomes." *Biometrics*, 42, 121-130.