

Medication Adherence in Cardiovascular Disease: Generalized Estimating Equations in SAS®

Erica Goodrich, Priority Health, Grand Rapids, MI

Daniel Sturgeon, Priority Health, Grand Rapids, MI

ABSTRACT

Medication non-adherence has been estimated to generate \$290 billion annually in medical costs (NEHI, 2009). Not only is this an extremely expensive dilemma, it has consequences including worsening conditions, increased hospitalizations and readmissions. Recent medical based evidence had led to an increased interest in investigating individuals' adherence to cardiovascular disease-related medications. Data from medical databases from a regional health plan of subjects on cardiovascular medications were sampled as subject matter. Binary data using a 70% possession ratio (PR) cutoff for determining medication adherence was analyzed over semi-quarterly time frames. Results of Generalized Estimating Equation (GEE) models are explained through SAS® output.

INTRODUCTION

Cardiovascular disease is the number one cause of death in the United States (CDC, 2011). Some health initiatives to help counter this risk and improve health outcomes include eating healthier, increasing exercise, and following stress management techniques. Literature also shows that adhering to medication regimens plays a major role in these types of events. Medication adherence, defined by the World Health Organization, is "the extent to which the patient follows medical instructions" (2011). The consequences of non-adherence include unnecessary morbidity and mortality, lower quality of life, and increased readmission rates, alongside increased hospital and health care costs (Ernst, 2007). In fact, New England Health Institute estimates that total potential savings from adherence and related disease management could be \$290 billion annually which is roughly equivalent to 13% of total health care spending in the US. Medication non-adherence is increasingly recognized as a very important issue due to its' significant health and economic consequences. Brown and Bussell estimated adherence to long-term therapy for chronic diseases at 50% in developed countries (2011).

There are many reasons for non-adherence that have been proposed and these can generally be categorized as provider-level (e.g., complex drug regimens, ineffective communication), patient-level (e.g., knowledge, motivation, finances, health literacy), and system-level (e.g., access to care, coordination of care, limited visits) (Brown & Bussell, 2011). As a health care system, being able to identify subjects who are most likely to be adherent to their medications are of interest. Some objectives of this analysis identifying traits related to individual adherence. By being able to identify subjects prospectively who may be less likely to adhere to medication therapy can assist with interventions and targeted outreach. The largest goal for this paper is to focus on the premise of when a Generalized Estimating Equations may be useful, and how it is used and interpreted in SAS®.

Subjects with cardiovascular disease were especially of interest, based on company-wide goals to focus on chronic disease alongside current medicine based literature. Currently, multiple drug classes related to Coronary Artery Disease (CAD) are being tracked which include ACE Inhibitors, AT(2) receptor antagonists, beta-blockers, calcium channel blockers, long acting nitrates, and statins. The Clinical Quality Improvement team has expressed interest in exploring subjects' medication complacence specifically on statins. Statins are a class of drugs used for lowering low-density lipoprotein (LDL) cholesterol which include products.

METHODS

Data from medical databases from a regional health plan of subjects on cardiovascular medications were sampled as subject matter. Claims data was used to evaluate filling patterns and computing a medication possession ratio (PR). This is the ratio of days supplied to days elapsed for a medication. The days supplied is computed by summing the days supplied field for a series of prescriptions, excluding the days supplied from the last prescription in the series (Steiner et. al 1998). It should be noted that the last prescription is used only to establish the end date. Days elapsed can be computed by subtracting the earliest fill date from the latest fill date. From there, any overlapping days in this range are subtracted. The PR assumes that the patient is, on average, refilling prescriptions at or close to the time that the previous prescription is exhausted.

PR is calculated on a semi-quarterly basis using prescriptions filled during the last 180 days of the 'reporting period.' To be included in these measures, a minimum of two prescriptions is required to compute PR. These calculations are

completed through vendor solutions clinical informatics software, Optum™ Impact Pro®. Impact Pro® then calculates at the patient level if a subject is compliant to their medication if they have a PR over 70% or not. This 70% value is often presented as a standard cutoff in many literature reviews and is displayed as a dichotomous 0/1 value. In the case of Figure 1, a person with a PR of 79.78% would be considered “1,” adherent with a PR above 70%. For our purposes this measure has two years worth of data collected. This provides four PR readings per subject.

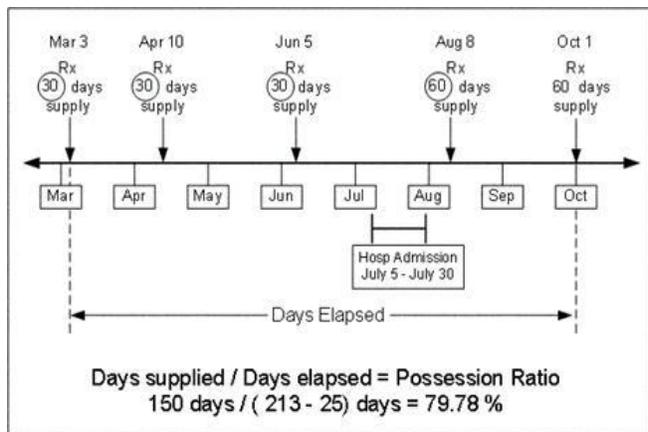


Figure 1. Possession Ratio calculation example by Impact Pro®.

INCLUSION CRITERIA

Subjects were identified who met the following criteria:

- Adult subjects over the age of 18
- Current subjects enrolled in Medicaid product
- Continuously enrolled over the past 24 months prior with medical benefits and pharmacy benefits
- Diagnosed with coronary artery disease or congestive heart failure based on Impact Pro®’s clinical definition
- Filled two or more statin class drug prescriptions in the past 180 days

From there, a simple random sample was taken using PROC SURVEYSELECT in which 1,547 subjects were selected for analysis.

VARIABLES OF INTEREST

Based on previous research and literature, a number of demographic and health related traits were identified to be possibly related to a person’s adherence to medication. For simplicity sake, a subsection of variables available were analyzed. Below shows the variables, the type of variable, and descriptive statistics.

Variable	Variable Type	Values	Descriptive Statistics
Age	Numeric	43-99	Mean =73.9, Median =73
Gender	Categorical	Male/Female	782/792
Depression	Categorical	Yes/No	306/1268
Charlson Co-morbidity Index(high)	Numeric	0-14	Mean = 1.44, Median = 3
Number of drugs	Numeric	1-15	Mean = 4.56, Median = 4
Previous adherence (time)	Categorical	Yes(1)/No(0)	--

Table 1. Variables of interest.

The selected variables above include the subject’s gender, whether or not they had been clinically diagnosed with depression, the number of drugs that the subject currently is taking in addition to a statin class drug, and four time points for the binary possession ratio value. Charlson co-morbidity index is a point based score used to predict 10 year mortality; however for our purposes we are using it as a proxy for co-morbidities. Note that only main effects are

tested in this model for simplicity and interpretation.

GENERALIZED ESTIMATING EQUATIONS

If there was only one time point available, a Generalized Linear Model (GLM) using logistic regression would be a suitable statistical test for this data, since medication adherence is captured as a binary variable in which a patient is adherent or non-adherent, and there are multiple independent variables of interest. However, in this example there are multiple periods collected on each individual. One assumption of GLMs is that the residuals are independent of one another, which is not the case in this study. This repeated time measurement provides additional information that we would like to add to the model.

Generalized Estimating Equations (GEE) were first discussed by Liang and Zeger (1986) as an extension of Generalized Linear Models. GEE models are used in instances where there are non-normal residuals with repeated measures. Since these data have repeated time points, the random component is described by the same variance function as in an independent case, but the covariance structure of the correlated measure must also be accounted for. To use GEE methodology, a link function and a distribution of the dependent variable must be also be defined. For an accessible and deeper explanation on GEE models, we recommend Liang and Zeger's original 1986 *Biometrika* "Longitudinal data analysis using generalized linear models" or Ziegler et al's *Biometrical Journal* titled "The Generalised Estimating Equations: An Annotated Bibliography".

For the current model of interest, there are 1,547 individual subjects, with four possession ratio measurements per subject. Each subject and their four measures are called a cluster. These subject's measures are correlated within possession ratios. To account for this correlated measure, a working correlation structure must be chosen and used. A working correlation matrix is a diagonal matrix which often is not known and needs to be estimated.

Often, prior to running a GEE model, a generalized linear model is performed. In this example, a logistic regression model could be created using PROC LOGISTIC, with similar syntax listed below without referencing the REPEATED line. In this instance, independence would be assumed, even though it is not the case to an initial estimate for the initial parameters (β). From here, residuals can be calculated and a working correlation matrix is estimated from these results. This can be an iterative process, in which a model is rerun and refit to continuously correct due to the correlation. After this is done, a GEE model can be fit. For the sake of interpreting results from this model (both significant and non-significant variables) this was not done for the scope of this paper.

While the most recent version of SAS/STAT Version 13.2 currently has an experimental procedure for Generalized Estimating Equations under PROC GEE, this version was not available to the authors at the time this paper was written. PROC GENMOD is the current established procedure for GEE models.

The SAS syntax needed for our model is as follows:

```
PROC GENMOD DATA=ds7 DESCENDING;  
  CLASS mbr_id sex cond_depression;  
  MODEL num = sex cond_depression age charlson time totaldrugs/LINK = LOGIT  
  DIST = BINOMIAL TYPE3;  
  REPEATED SUBJECT = mbr_id / TYPE = EXCH COVB CORRW;  
RUN;
```

Starting at the beginning of the PROC GENMOD, the dataset of interest is noted through the DATA= command. The DESCENDING option is used to change the default model's probability calculations. Versions after SAS Version 8.1 default to modeling the dependent variable outcome being equal to 0. By using DESCENDING, the dependent variable outcome of 1 is being modeled. This is critical to note when interpreting your results that you are reporting on the proper outcome and making the correct comparisons.

The CLASS statement is used for categorical variables which have multiple levels. In the instance above, depression status, sex, and subject are included. The MODEL statement is used to specify the response variable along with the dependent variables. In this example, num is being used as the response variable. This is the binary value for whether or not a subject has a possession ratio above 80%. The LINK and DIST commands are being used to identify that the link structure of interest is a logit, and the distribution is binomial. Depending on the data available and the study of interest Generalized Estimating Equations can be used with different distributions or links. Some variance functions include a normally-distributed response with an identity link, and a poisson response with a log link. PROC GENMOD recognizes that GEE is being invoked through using the REPEATED statement. In this instance, we are noting that the cluster or subject is being called mbr_id. TYPE3 provides Type 3 analysis statistics

output, discussed further below. Options on the REPEATED line include TYPE=, COVB and CORRW. The COVB displays the estimated regression parameter covariance matrix, and CORRW provides the estimated correlation matrix output. The TYPE= option is used for using the correlation structure of interest. A working correlation structure is needed as a correction of within subject correlation. The exchangeable working correlation (EXCH) was chosen in this project. Exchangeable working correlation assumes that all distinct PR measures of a cluster, y_{ij} , are equally correlated with each other and the total variation. An example of what the correlation matrix would look like can be viewed below:

$$\text{Corr}(y_{ij}, y_{i,j'}) = \begin{cases} 1 & j = j' \\ \alpha & j \neq j' \end{cases} \quad \mathbf{R} = \begin{bmatrix} 1 & \alpha & \alpha & \alpha \\ \alpha & 1 & \alpha & \alpha \\ \alpha & \alpha & 1 & \alpha \\ \alpha & \alpha & \alpha & 1 \end{bmatrix}$$

The SAS output for working correlation matrix shows this in action.

SAS OUTPUT RESULTS

Here, all PROC GENMOD output will be explained in further detail starting at the beginning of the output. Basic model information is provided. This provides information on which dataset is chosen, which distribution that was chosen, which link function was chosen, and the dependent variable being used in the model. These all have been specified in the code listed above.

Model Information	
Data Set	WORK.DS7
Distribution	Binomial
Link Function	Logit
Dependent Variable	num

The next output provided shows class level information. This lists the variables in the class statement, and the level count and specification of what those levels are. For this analysis, there were 1,574 different subjects in this analysis. Usually these are shown in the class information, along with any other categorical variables in the class statement. For this analysis gender (coded as sex) and depression status (coded as cond_depression) are the only being used.

Class Level Information		
Class	Levels	Values
MBR_ID	1574	***
SEX	2	F M
COND_DEPRESSION	2	N Y

For the response profile, this shows the count of observations for each of the binary options of the dependent variable. This accounts for the 1,574 subjects with four time points each.

Response Profile		
Ordered Value	num	Total Frequency
1	1	6126
2	0	170

SAS also produces the note: "PROC GENMOD is modeling the probability that num='1'."

SAS also provides parameter information provides index numbers and information needed for the parameters of interest. This table is generated for this example due to the use of COVB in the option statement. By default, class variables are organized alphabetically.

Parameter Information			
Parameter	Effect	SEX	COND_DEPRESSION
Prm1	Intercept		
Prm2	SEX	F	
Prm3	SEX	M	
Prm4	AGE		
Prm5	COND_DEPRESSION		N
Prm6	COND_DEPRESSION		Y
Prm7	CHARLSON		
Prm8	time		
Prm9	TOTALDRUGS		

Covariance Matrix (Model-Based)							
	Prm1	Prm2	Prm4	Prm5	Prm7	Prm8	Prm9
Prm1	1.54915	-0.01255	-0.02041	-0.02155	0.006072	-0.004927	-0.01209
Prm2	-0.01255	0.06566	-0.000429	0.01331	0.0004962	0.0000517	0.001597
Prm4	-0.02041	-0.000429	0.0002945	-0.000564	-0.000147	4.5142E-6	8.8968E-6
Prm5	-0.02155	0.01331	-0.000564	0.07513	0.002837	0.0000980	-0.000260
Prm7	0.006072	0.0004962	-0.000147	0.002837	0.002389	-0.000051	-0.000850
Prm8	-0.004927	0.0000517	4.5142E-6	0.0000980	-0.000051	0.002550	0.0000114
Prm9	-0.01209	0.001597	8.8968E-6	-0.000260	-0.000850	0.0000114	0.003450

Covariance Matrix (Empirical)							
	Prm1	Prm2	Prm4	Prm5	Prm7	Prm8	Prm9
Prm1	1.08855	-0.01990	-0.01321	-0.02442	-0.000581	-0.002844	-0.02256
Prm2	-0.01990	0.08060	-0.000905	0.04180	0.002154	0.001488	0.004124
Prm4	-0.01321	-0.000905	0.0002050	-0.001081	-0.000081	-0.000067	-0.000047
Prm5	-0.02442	0.04180	-0.001081	0.09109	0.001937	0.002407	0.003482
Prm7	-0.000581	0.002154	-0.000081	0.001937	0.001990	-0.000079	-0.000042
Prm8	-0.002844	0.001488	-0.000067	0.002407	-0.000079	0.002517	0.0002851
Prm9	-0.02256	0.004124	-0.000047	0.003482	-0.000042	0.0002851	0.005629

Working Correlation Matrix				
	Col1	Col2	Col3	Col4
Row1	1.0000	0.5207	0.5207	0.5207
Row2	0.5207	1.0000	0.5207	0.5207
Row3	0.5207	0.5207	1.0000	0.5207
Row4	0.5207	0.5207	0.5207	1.0000

Exchangeable Working Correlation	
Correlation	0.5207357978

This output provides information on the exchangeable covariance structure estimates and then the working correlation matrix. The first window for the Model-based covariance matrix is also known as the naïve variance estimator. This is consistent when the mean model and the covariance model are correctly identified. The empirical based covariance matrix is also known as the sandwich or robust/semi-robust variance estimator. This is consistent when the mean model is correctly specified and there is no missing data.

GEE Fit Criteria	
QIC	1520.4423
QICu	1500.3459

The GEE Fit criteria is used as a model fit criteria. QIC statistics are used for comparing different working correlation structures. In some instances a model may be run with different working correlation structures. The model with the smallest QIC values is best.

Score Statistics For Type 3 GEE Analysis			
Source	DF	Chi-Square	Pr > ChiSq
SEX	1	2.01	0.1563
AGE	1	5.88	0.0153
COND_DEPRESSION	1	3.38	0.0661
CHARLSON	1	9.13	0.0025
time	1	15.93	<.0001
TOTALDRUGS	1	0.29	0.5881

The Type 3 Analysis provides Wald χ^2 test statistic for testing the significance of each variable, and their respective degrees of freedom and p-values. From this output, there are significant values for age, Charlson co-morbidity index, and time and a marginally significant value for depression status.

Analysis Of GEE Parameter Estimates						
Empirical Standard Error Estimates						
Parameter		Estimate	Standard Error	95% Confidence Limits		Z Pr > Z
Intercept		0.1063	1.0433	-1.9386	2.1512	0.10 0.9189
SEX	F	0.3976	0.2839	-0.1589	0.9540	1.40 0.1614
SEX	M	0.0000	0.0000	0.0000	0.0000	. .
AGE		0.0358	0.0143	0.0077	0.0638	2.50 0.0125
COND_DEPRESSION	N	0.6209	0.3018	0.0294	1.2125	2.06 0.0397
COND_DEPRESSION	Y	0.0000	0.0000	0.0000	0.0000	. .
CHARLSON		-0.2171	0.0446	-0.3045	-0.1296	-4.87 <.0001
time		0.2038	0.0502	0.1054	0.3021	4.06 <.0001
TOTALDRUGS		0.0385	0.0750	-0.1086	0.1855	0.51 0.6080

The output titled “Analysis of GEE Parameter Estimates” is used for viewing the results of the model fit. The variables total number of drugs and sex are not significant which would make them variables which could be removed from the model since they are not adding any additional information to the model. For the sake of simplicity and the keeping the original variables of interest in the model, a reduced model will not be discussed. The lack of significance could have also been addressed with the initial logistic regression that was performed. Significant predictors in medication adherence were age, whether or not a subject had depression, Charlson co-morbidity score, and time (previous PR outcomes).

INTERPRETATION

Output from a GEE analysis using a logit link function and binomial distribution which are interpreted by using the exponential value of the estimate. Also, since this is a larger sample size, it is important to note confidence intervals, odds ratios (OR), or even effect sizes. P-values may state that there is statistical significance; however this may not be helpful for interpreting clinical significance or even a meaningful significance (Chen et al).

Age is a significant predictor ($Z = 2.5$, $p\text{-value} = 0.0125$) with the β estimate of 0.0358. For an increase in one year of age, a subject is 3.6% ($= e^{0.0358}$) more likely to be adherent than a subject a year younger when holding all other variables equal. An example would be comparing a 55 year old to a 56 year old subject on average. This can viewed in a multiplicative manner also. Taking this interpretation further, when comparing a 55 year old and 65 year old subject, an increase in 10 years is roughly 36.5% more likely to be adherent on average. It should be noted that this is a very small OR value, so it has a small effect size. For interpreting the confidence interval, we could say we are 95% confident that an increase in one year of age, a subject is between .7% ($= e^{0.0077}$) and 6.5% ($= e^{0.0638}$) more likely to be adherent than a subject a year younger.

Depression status is also significant ($Z = 2.06$, $p\text{-value} = 0.0397$) with an estimate of 0.6209. On average, holding all other variables equal, the odds of a subject without a clinical depression diagnosis compared to a subject with a clinical depression diagnosis is 86% ($= e^{0.6209}$) more likely to be adherent to their medication.

Charlson co-morbidity Index also are significant ($Z = -4.87$, $p\text{-value} < .0001$) and have an estimate of -0.2171. A subject is 0.8 ($= e^{-0.2171}$) times as likely to be adherent for every increase in one Charlson co-morbidity index point. Take into account that this Z-statistic and estimate are both negative values. This is important to note for proper interpretation.

The odds of medication adherence with a PR above 70% in later PR outcome periods is 23% ($= e^{0.2038}$) higher than those in earlier periods ($Z = 4.06$, $p\text{-value} < .0001$).

CONCLUSION

Until somewhat recently performing large scale generalized estimating equations would have been seen as too resource heavy to warrant using them as a method of analysis. Thanks to an increase in computer power, and procedures like PROC GENMOD and soon PROC GEE there has been an increase in the use of this method in the biomedical sciences for longitudinal data. The healthcare industry specifically has been a rich source of longitudinal data; however advanced statistical analysis has not always been adapted or embraced. By using methods such as GEE, we have been able to use proper and informative statistical methodology for further insight. For our purposes we wanted to analyze medication adherence over time, and GEE provided the necessary information to make a positive impact in the lives of our members who may have had worse outcomes without an intervention.

REFERENCES

- Brown, M. T., & Bussell, J. K. (2011). Medication Adherence: WHO Cares? *Mayo Clinic Proceedings*, 86(4), 304–314. doi:10.4065/mcp.2010.0575
- CDC (2011). Million Hearts: strategies to reduce the prevalence of leading cardiovascular disease risk factors. United States, *MMWR*2011; 60(36):1248–51.
- Chen H, Cohen P, Chen S (2010). How Big is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. *Communications in Statistics - Simulation and Computation*, 39: 860-864.
- Cramer J.A., Roy A., Burrell A., Fairchild C.J., Fuldeore M.J., Ollendorf D.A., Wong P.K. (2008) Medication Compliance and Persistence: Terminology And Definitions. *Value in Health* 11.1: 44-47.
- EBM Symmetry Impact Pro® Hypertension Algorithm Guide.
- Ernst F.R., Grizzle A.J. (2001). Drug-related morbidity and mortality: updating the cost-of-illness model. *Journal of the American Pharmaceutical Association*, 41(2):192–199. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11297331>.
- Goodrich, E.L. (2014). Exploration of Repeated Measures Logistic Regression for Medication Adherence. Joint Research Conference. Seattle, WA
- International Society of Pharmacoeconomics and Outcomes Research (ISPOR) Medication Compliance and Persistence Special Interest Group (2009). <http://www.ispor.org/sigs/medication.asp>
- Leslie, R.S. (2009). "Calculating Medication Compliance, Adherence, and Persistence in Administrative Pharmacy Claims Databases" <http://www.wuss.org/proceedings08/08WUSS%20Proceedings/papers/anl/anl09.pdf>.
- Kravitz R.L., Hays R.D., Sherbourne C.D., DiMatteo M.R., Rogers W.H., Ordway L., Greenfield S. (1993). Recall of recommendations and adherence to advice among patients with chronic medical conditions.
- Liang K. and Zeger S. (1986). "Longitudinal data analysis using generalized linear models". *Biometrika* 73 (1): 13–22
- Sabaté, E., & World Health Organization (2003). *Adherence to long-term therapies: Evidence for action*. Geneva: World Health Organization.
- Stokes, M., Davis, C., and Koch, G. (2012). *Categorical data analysis using the SAS system*. 3rd ed. Cary, NC: SAS Institute.
- Tamburrino, M. B., Nagel, R. W., Chahal, M. K., & Lynch, D. J. (2009). Antidepressant Medication Adherence: A Study of Primary Care Patients. *Primary Care Companion to The Journal of Clinical Psychiatry*, 11(5), 205–211. doi:10.4088/PCC.08m00694
- Steiner J.F., Koepsell T.D., Fihn S.D., & Inui T.S. A general method of compliance assessment using centralized pharmacy records. *Med Care*. 1988 Aug;26(8):814–823.
- Ziegler, A., Kastner, C. and Blettner, M. (1998). The generalized estimating equations: an annotated bibliography, *Biometrical Journal*, 40 (2), 115-139.

APPENDIX

Impact Pro®'s disease condition could be confirmed through disease registry, professional encounter. Code set, professional supervision, or a facility event for admission, emergency room, or outpatient surgery. The subjects who had two or more fills of a statin class drug in the past 180 days were included in this study.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Erica Goodrich
Enterprise: Priority Health
Address: 1231 East Beltline Ave NE
City, State ZIP: Grand Rapids, MI 49525
Work Phone: 616-464-8142
Fax: 616-942-0024
E-mail: erica.goodrich@priorityhealth.com

Name: Daniel Sturgeon
Enterprise: Priority Health
Address: 1231 East Beltline Ave NE
City, State ZIP: Grand Rapids, MI 49525
Work Phone: 616-575-5740
Fax: 616-942-0024
E-mail: daniel.sturgeon@priorityhealth.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.