Paper 1335-2015

Using GLIMMIX and GENMOD Procedures to Analyze Longitudinal Data from a Department of Veterans Affairs Multisite Randomized Controlled Trial

Abbas S. Tavakoli, DrPH, MPH, ME[1]; Nikki R. Wooten, PhD, LISW-CP[2,3], Marlene B. Al-Barwani, BS[2]; Selina H. McKinney, PhD, APRN-BC, P/MHNP-BC[1], Sue E. Levkoff, ScD[2]

[1]University of South Carolina, College of Nursing
[2]University of South Carolina, College of Social Work
[3]William Jennings Bryan Dorn Veterans Administration Hospital

Abstract

Many SAS® procedures can be used to analyze longitudinal data. This study employed a multisite randomized controlled trial design to demonstrate the effectiveness of two SAS procedures, GLIMMIX and GENMOD, to analyze longitudinal data from five Department of Veteran Affairs Medical Centers (VAMCs). Older male veterans (n = 1222) seen in VAMC primary care clinics were randomly assigned to two behavioral health models, integrated (n = 605) and enhanced referral (n = 617). Data were collected at baseline, and 3, 6, and 12 month follow-up. A mixed-effects repeated measures model was used to examine the dependent variable, problem drinking, which was defined as count and dichotomous from baseline to 12 month follow-up. Sociodemographics and depressive symptoms were included as covariates. First, bivariate analyses included general linear model and chi-square tests to examine covariates by group and group by problem drinking outcomes. All significant covariates were included in the GLIMMIX and GENMOD models. Then, multivariate analysis included mixed models with Generalized Estimation Equations (GEEs). The effect of group, time, and the interaction effect of group by time were examined after controlling for covariates. Multivariate results were inconsistent for GLIMMIX and GENMOD using Lognormal, Gaussian, Weibull, and Gamma distributions. SAS is a powerful statistical program in data analyses for longitudinal study.

Keywords: SAS, mixed models, GLIMMIX, GENMOD, longitudinal data

Introduction

When analyzing longitudinal data, it is challenging for new investigators to determine what type of statistical procedures to use and which statistical software is most powerful and efficient (Singer et al 2003). SAS is a powerful statistical software program and provides multiple efficient procedures for investigators to analyze longitudinal data. The GLIMMIX procedure performs a variety of mixed linear models, and fits estimation and statistical inferences for various study designs including longitudinal and cohort designs (Little et al 1996). The GENMOD procedure in SAS is used to perform general linear models as well as nonlinear and complex models including log linear, logistic, or count models. GENMOD also accounts for correlated outcomes that can be expected in longitudinal studies when outcomes are measured over time. The GLIMMIX procedure extends the GENMOD procedure by estimating mixed models with data from non-normal distributions. SAS provides practical and efficient ways to analyze longitudinal data collected using complex study designs.

Purpose

The purpose of this paper is to compare GLIMMIX and GENMOD procedures using SAS to analyze longitudinal data with multiple distributions.

Background

This paper uses longitudinal data to conduct a secondary analysis of the Primary Care Research in Substance Abuse and Mental Health for Elderly (PRISM-E) study, a multisite randomized controlled trial comparing integrated and enhanced referral behavioral health models among primary care patients aged 65 and older with mental health and substance use problems seen in civilian hospitals and Department of Veteran Affairs Medical Centers (VAMCs). Data analyzed for this study were collected at five VAMCs (Chicago, Dartmouth, Madison, Miami, and Philadelphia) at which patients aged 65 and older were randomized to integrated and enhanced referral behavioral health (BH) models. Data were collected at baseline, and 3, 6, and 12 month follow-up. Integrated and enhanced referral BH models were available at the five VAMC sites for six months prior to patient randomization. The integrated model included: (1) mental health and/or substance abuse (MHSA) services co-located within primary care clinics, (2) communication between the Primary Care Physician (PCP) and MHSA provider about the behavioral health evaluation and treatment plan, and (3) availability of brief alcohol treatment interventions for at-risk drinkers. The enhanced model included: (1) MHSA evaluation and treatment by licensed BH providers at another location, (2) coordinated follow up with primary care clinics if the patient missed the first scheduled appointment, and (3) transportation assistance. A more detailed description of PRISM-E is explained in Levkoff et al. (2004) and Zanjani et al. (2008).

Alcohol misuse may be categorized as at-risk drinking, problem drinking, or alcohol abuse or dependence (Blow & Barry, 2012; NIAAA, 2008; Saitz, 2005). At-risk drinking exceeds alcohol use guidelines established by NIAAA and increases potential for future health problems. NIAAA recommends men limit alcohol intake to 14 drinks per week, women and healthy older men who do not take medication to 7 drinks per week or no more than one drink per day for individuals who combine alcohol use with certain prescription medications (e.g. benzodiazepines; NIAAA, 2008; Saitz, 2005). Problem drinking also exceeds these guidelines and has led to diminished physical, social, or emotional functioning. Alcohol abuse and dependence are characterized by continued alcohol use despite these adverse consequences.

Problem drinking was the primary outcome for the current study and was indicated by endorsement of more than 3 items on the Short Michigan Alcohol Screening Test-Geriatric Version (SMAST-G). SMAST-G assesses alcohol-related problems in physical, emotional, and social contexts. SMAST-G scores of 2 or more indicate the possibility of problem drinking. Higher scores indicate increased level of severity. For the purposes of this study, a conservative approach was taken to identify older male veterans with a high probability of problem drinking behavior. The sample included older male veterans (n=1222) aged 65 to 93 (M_{age} = 73.8) randomized to an integrated or enhanced referral BH treatment model. Problem drinking was operationalized as SMAST-G score of 3 or higher (y/n) and total SMAST-G score (continuous). Potential covariates included age, race, gender, education, employment status, Medicaid, Medicare, VA benefit, VA site, and depressive symptoms. A detailed description of the sample and the prevalence of problem drinking among these older male veterans is explained in Wooten et al. (2014).

Methods

Data were merged by unique participant ID. PROC MEAN and PROC FREQ were used to describe prevalence of problem drinking in each treatment group at four time points. T=tests, chi-square, and generalized linear models (GLM) were performed to examine covariates. PROC GLIMMIX and GENMOD were used to analyze longitudinal data. Only significant covariates (VA site, race, depressive symptoms, assigned BH treatment group, and time) were included in multivariate models. Problem drinking as a count variable was examined with different distributions using GLIMMIX. The group, time, and group by time interaction effects were examined for all models after controlling for covariates. All data analyses were performed using SAS/STAT® version 9.4 (SAS, 2013).

Results

Table 1. Measure of center and dispersion for problem drinking by group and time

Behavioral Health Groups	Time	N	Mean	*SD*	Minimum	Maximum
Integrated	1	217	2.53	2.27	0.00	10.00
	2	178	2.35	2.32	0.00	10.00
	3	178	2.19	2.12	0.00	10.00
	4	116	2.28	2.06	0.00	10.00
Enhanced Referral	1	217	2.82	2.26	0.00	9.00
	2	177	2.33	2.16	0.00	10.00
	3	184	2.37	2.03	0.00	8.00
	4	129	2.26	2.09	0.00	8.00

Table 1 shows descriptive statistics for problem drinking by group and time. For the integrated group, the mean of problem drinking decreased from baseline to time 3 and increased slightly from time 3 to time 4. For the enhanced referral group, the mean of risking drinking decreased from time 1 to time 2, but increased slightly from time 2 to time 3, and decreased from time 3 to time 4. Table 2 indicated frequency distribution of the dichotomous problem drinking outcome by time for each group. The time by problem drinking outcome for the enhanced referral group was significant (p = .0351).

Table 2: Frequency distribution of problem drinking by time for integrated behavioral health group

Problem Drinking by Time					
Integrated Behavioral Health Model					
Problem Drinking	Time				
Frequency Percent Row Pct Col Pct	1	2	3	4	Total
No	124 18.00 29.59 57.14	114 16.55 27.21 64.04	111 16.11 26.49 62.36	70 10.16 16.71 60.34	419 60.81
Yes	93 13.50 34.44 42.86	64 9.29 23.70 35.96	67 9.72 24.81 37.64	46 6.68 17.04 39.66	270 39.19
Total	217 31.49	178 25.83	178 25.83	116 16.84	689 100.00
Frequency Missing = 1731					

Statistic	DF	Value	*P*
Chi-Square	3	2.1961	0.5327
Likelihood Ratio Chi-Square	3	2.1930	0.5333
Mantel-Haenszel Chi-Square	1	0.4829	0.4871
Phi Coefficient		0.0565	
Contingency Coefficient		0.0564	
Cramer's V		0.0565	

Table 3: Frequency distribution of problem drinking by time for referral behavioral health group

Problem Drinking by Time					
Referral Behavioral Health Model					
Problem Drinking	**Time**				
Frequency Percent Row Pct Col Pct	**1**	**2**	**3**	**4**	**Total**
No	110 15.56 27.64 50.69	101 14.29 25.38 57.06	101 14.29 25.38 54.89	86 12.16 21.61 66.67	398 56.29
Yes	107 15.13 34.63 49.31	76 10.75 24.60 42.94	83 11.74 26.86 45.11	43 6.08 13.92 33.33	309 43.71
Total	217 30.69	177 25.04	184 26.03	129 18.25	707 100.00
Frequency Missing = 1761					

Statistic	DF	Value	*P*
Chi-Square	3	8.5993	**0.0351**
Likelihood Ratio Chi-Square	3	8.7259	0.0332
Mantel-Haenszel Chi-Square	1	6.3394	0.0118
Phi Coefficient		0.1103	
Contingency Coefficient		0.1096	
Cramer's V		0.1103	

Tables 2 and 3 indicate frequency distribution of the dichotomous problem drinking outcome by time for each group. The time by problem drinking outcome for the enhanced referral group was significant (*p* = .0351).

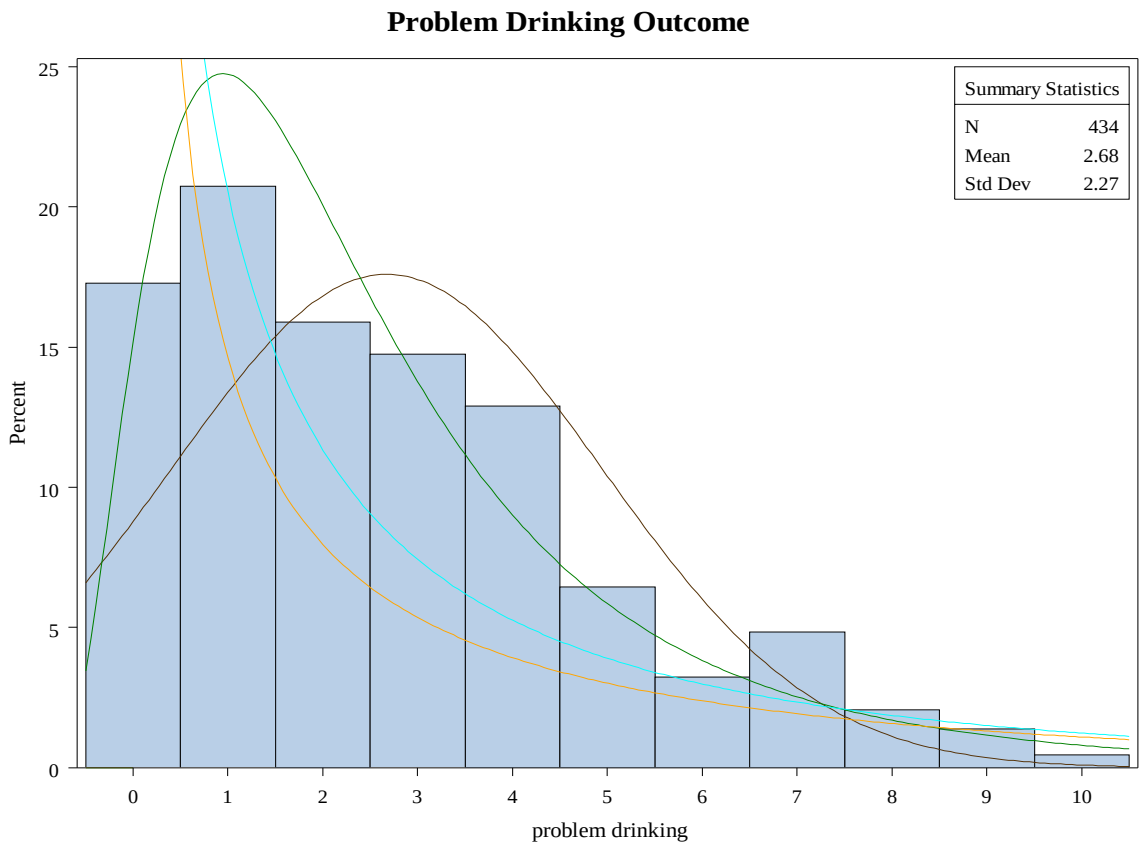

Figure. Problem drinking distribution for baseline.

Note. Lognormal Gaussian Weibull Gamma

Figure 1 showed problem drinking distribution is not normal. The figure included Gaussian, Lognormal, Weibull and Gamma distributions.

Table 4. Type III Tests output for GLIMMIX procedure with different distribution

Type III Tests of Fixed Effects											
Effect	Num DF	Gaussian F Value	Gaussian P-Value	Lognormal F Value	Lognormal P-Value	Weibull F Value	Weibull P-Value	Gamma F Value	Gamma P-Value	Binomial F Value	Binomial P-Value
VA SITE	4	3.27	**0.0119**	7.38	**<.0001**	3.79	**0.0050**	7.46	**<.0001**	4.78	**0.0009**
RACE[a]	2	4.33	**0.0139**	1.65	0.1928	4.07	**0.0178**	1.79	0.1686	1.94	0.1455
CESD_CUT[b]	1	58.90	**<.0001**	50.74	**<.0001**	53.30	**<.0001**	49.90	**<.0001**	32.03	**<.0001**
BH GROUP	1	4/16	**0.0420**	2.99	0.0845	4.56	**0.0334**	3.44	0.0644	4.30	**0.0390**
TIME	3	6.05	**0.0004**	2.14	0.0942	5.46	**0.0010**	3.76	**0.0107**	2.27	0.0794
Time*Assign	3	1.54	0.2033	0.92	0.4312	1.25	0.2896	0.88	0.4506	2.42	0.0650

Note. VA = Veterans Affairs; CESD = Center for Epidemiologic Studies of Depression Scale
[a]RACE (1=White, 2=Black, 3=Other) [b]CESD_CUT: Depressive symptoms (1=Yes, 0=No)

Table 4 shows Type III tests from GLIMMIX output with different distributions for problem drinking outcome as count and dichotomous. Between group effects were significant for Gaussian (p=.0420), Weibull (*p* =.0334), and binomial (*p* =.0390) after controlling for site, race, and depression. Time effects were significant with all distributions except binominal (p=.0794). The within-between group and time effects were not significant for all distributions. The result showed site and depression were significant among all distributions. However, race was only significant for Gaussian and Weibull distribution.

Table 5. Parameter estimate and Type III Tests output for GENMOD procedure

Analysis Of GEE Parameter Estimates						
Empirical Standard Error Estimates						
Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	1.3353	0.4092	0.5334	2.1373	3.26	0.0011
Chicago VA (ref = Philadelphia)	0.5206	0.3671	-0.1989	1.2401	1.42	0.1562
Dartmouth/WRJ VA (ref = Philadelphia)	0.1534	0.3174	-0.4686	0.7755	0.48	0.6288
Madison VA (ref = Philadelphia)	-0.7970	0.3212	-1.4267	-0.1674	-2.48	**0.0131**
Miami VA (ref = Philadelphia)	0.3691	0.3606	-0.3377	1.0758	1.02	0.3060
African American (ref = White)	0.5794	0.3276	-0.0627	1.2214	1.77	0.0770
Other (ref = White)	0.2880	0.7378	-1.1581	1.7342	0.39	0.6962
CESD (ref = no)	-1.6610	0.3158	-2.2798	-1.0421	-5.26	**<.0001**
BH Model (ref = integrated)	-0.5151	0.2431	-0.9915	-0.0387	-2.12	**0.0341**
Time (ref = 2)	-0.2580	0.1661	-0.5835	0.0676	-1.55	0.1204

Analysis Of GEE Parameter Estimates						
Empirical Standard Error Estimates						
Parameter	**Estimate**	**Standard Error**	**95% Confidence Limits**		**Z**	**Pr > |Z|**
Time (ref = 3)	-0.0626	0.1707	-0.3971	0.2719	-0.37	0.7137
Time (ref = 4)	-0.6894	0.2083	-1.0976	-0.2812	-3.31	**0.0009**
Time x Assign (ref = 2, integrated)	-0.0596	0.2605	-0.5701	0.4509	-0.23	0.8191
Time x Assign (ref = 3, integrated)	-0.1099	0.2612	-0.6219	0.4021	-0.42	0.6740
Time x Assign (ref = 4, integrated)	0.6943	0.3141	0.0787	1.3100	2.21	**0.0271**

Score Statistics For Type 3 GEE Analysis			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
VA SITE	4	19.38	**0.0007**
RACE	2	3.01	0.2221
CESD_CUT	1	23.14	**<.0001**
ASSIGN	1	4.47	**0.0344**
TIME	3	11.39	**0.0098**
TIME x GROUP	3	6.91	0.0749

Table 5 indicates that both between effect (group) and within effect (time) are significant. However, the interaction effect was not significant (p=.0749). Site and depression were significant in the model.

Table 6. Parameter estimate and Odds Ratio from GENMOD procedure

Contrast Estimate Results										
Label	**Mean Estimate**	**Mean Confidence Limits**		**L'Beta Estimate**	**Standard Error**	**Alpha**	**L'Beta Confidence Limits**		**Chi-Square**	**Pr > ChiSq**
INTEGRATED	0.3740	0.2706	0.4903	-0.5151	0.2431	0.05	-0.9915	-0.0387	4.49	**0.0341**
Exp (integrated)				0.5974	0.1452	0.05	0.3710	0.9620		
CESD (ref = yes)	0.8404	0.7393	0.9072	1.6610	0.3158	0.05	1.0421	2.2798	27.67	**<.0001**
Exp (CESD, yes)				5.2643	1.6622	0.05	2.8352	9.7749		
BLACK	0.6409	0.4843	0.7723	0.5794	0.3276	0.05	-0.0627	1.2214	3.13	0.0770
Exp(black)				1.7849	0.5847	0.05	0.9392	3.3920		
OTHER	0.5715	0.2390	0.8499	0.2880	0.7378	0.05	-1.1581	1.7342	0.15	0.6962
Exp(other)				1.3338	0.9841	0.05	0.3141	5.6641		
TIME 3	0.4359	0.3581	0.5169	-0.2580	0.1661	0.05	-0.5835	0.0676	2.41	0.1204
Exp(time 3)				0.7726	0.1283	0.05	0.5580	1.0699		
TIME 6	0.4844	0.4020	0.5676	-0.0626	0.1707	0.05	-0.3971	0.2719	0.13	0.7137
Exp(time 6)				0.9393	0.1603	0.05	0.6723	1.3124		
TIME 12	0.3342	0.2502	0.4302	-0.6894	0.2083	0.05	-1.0976	-0.2812	10.96	**0.0009**
Exp(time 12)				0.5019	0.1045	0.05	0.3337	0.7549		

Note: CESD = Center for Epidemiologic Studies of Depression Scale; Exp = Odds Ratio

Table 6 shows the parameter estimate and odds ratio (OR) for each parameter. The results indicated that those who were assigned to the integrated behavioral health treatment were .59 times (95% CI: 0.37-0.96) less likely to be risky drinkers compared to those assigned to enhanced referral behavioral health referral treatment after adjusting for correlated outcome data and controlling for race, depression, and site.

Conclusion

Our data indicated GLIMMIX and GENMOD provide slightly different results. The GLIMMIX procedure easily examined different distributions for our outcome, problem drinking, defined as more than one drink daily and a SMAST-G score of 3 or more. The GENMOD procedure calculates measure of association (odds ratio) for dichotomous outcomes. The GENMOD is a powerful procedure in SAS to perform general linear models as well as general estimating equations when the data are correlated such as with longitudinal data. It is good idea to examine different methods when the outcome is not normally distributed. SAS is one of the most powerful statistical programs for the analysis of longitudinal data.

References

Blow, F. C. & Barry, K. L. (2012). Alcohol and substance misuse in older adults. *Current Psychiatry Reports, 14*, 310-319.

Levkoff, S. E., Chen, H., Coakley, E., McDonel Herr, E. C., Oslin, D. W., Katz, I., … Ware, J. H. (2004). Design and sample characteristics of the PRISM-E multisite randomized trial to improve behavioral health care for the elderly. *Journal of Aging and Health, 16,* 3-27. doi: 10.1177/0898264303260390

Littell, R. C., Miliken, G., Stroup, W., & Wolfinger, R. (1996). SAS system for mixed models. SAS Institute Inc. *Cary, NC.*

NIAAA. (2008). *Older Adults*. Retrieved from National Institute of Alcohol Abuse and Alcoholism: http://www.niaaa.nih.gov/alcohol-health/special-populations-co-occurring-disorders/older-adult

Saitz, R. (2005). Unhealthy alcohol use. *The New England Journal of Medicine, 352,* 596-607.

Singer, J. D. & Willett, J. B. *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press; 2003.

Wooten, N. R., Tavakoli, A. S., Al-Barwani, M. B., Thomas, N. A., McKinney, S. H., Scheyett, A. M., … Levkoff, S. E. (2014). *A multisite randomized controlled trial of VA integrated and enhanced referral behavioral health models on alcohol misuse among older male veterans*. Manuscript in preparation.

Zanjani, F., Mavandadi, S., TenHave, T., Katz, I., Durai, N. B., Krahn, D., … Oslin, D. W. (2008). Longitudinal course of substance treatment benefits in older male veteran at-risk drinkers. *Journal of Gerontology Series A: Biological Sciences and Medical Sciences, 63A* (1): 98-106.

SAS Institute Incorporated. (2013). *SAS for Windows 9.4*. Cary, NC: SAS Institute Inc.

Contact Information

Your comments and questions are valued and encouraged. Please contact the authors at:

Abbas S. Tavakoli, DrPH, MPH, ME
College of Nursing
University of South Carolina
1601 Greene Street
Columbia, SC 29208-4001
Fax: (803) 777-5561
E-mail: abbas.tavakoli@sc.edu

Nikki R. Wooten, PhD, LISW-CP
College of Social Work
University of South Carolina
902 Sumter Street
Columbia, SC 29208
Email: nwooten@sc.edu

Marlene B. Al-Barwani, BS
College of Social Work
University of South Carolina
902 Sumter Street
Columbia, SC 29208
Email: bates2@mailbox.sc.edu

Selina H. McKinney, PhD
College of Nursing
Univeristy of South Carolina
1601 Greene Street
Columbia, SC 29208-4001
Email: hunts@mailbox.sc.edu

Sue E. Levkoff, ScD
College of Social Work
University of South Carolina
902 Sumter Street
Columbia, SC 29208
Email: slevkoff@mailbox.sc.edu

Attachment A

SAS Syntax

```
ods rtf;
ods listing close;
proc means data=four maxdec=2 ;
    class assign time;
      var tcsmas ;
    title ' means  / by assign and time';
    title2 'PRISM study';
   run;

proc freq data =four;
tables tcsmas ;
title ' Frequency tables  /   '    ;
           title2 'PRISM study';
    run;

proc freq data =four;
tables Assign* time *tcsmasg /chisq;
title ' Frequency tables  /   '    ;
           title2 'PRISM study';
    run;
ods rtf close;
ods listing;
quit;
run;

ods rtf;
ods listing close;


symbol;
goptions ftext= ctext= htext=;

 proc univariate data=four;
  var tcsmas;
  where time=1;
  histogram / midpoints= 0 to 10 by 1
            normal
            lognormal ( zeta=est sigma=est theta=est
                     w=1 color=GREEN noprint)
            weibull ( sigma=est theta=est w=1
                      color=CYAN noprint )
            gamma ( sigma=est theta=est w=1
                      color=orange noprint )
            vaxis   = axis1
            name    = 'MyHist';
  inset n mean(5.2) std='Std Dev'(5.2)  / pos = ne  header = 'Summary Statistics';
  axis1 label=(a=90 r=0);
title ' Problem Drinking Outcome ';
```

```
run;

ods rtf close;
ods listing;
quit;
run;

ods rtf;
ods listing close;
%macro gmix (a,b,c,d,e,f);
proc glimmix data=four;
        class pid time &a ;
    model &b= &c  /dist=&d  solution ;
    random time / subject= pid(assign) type=&e rside;
           title ' Glimmix/ Poisson model' &f ;
          title2 'PRISM study';
 run;

%mend gmix;
%gmix (assign siteid racegb  cesd_cut ,tcsmas, siteid racegb cesd_cut assign time assign*time
,normal,cs,Drinking risk/normal);
%gmix (assign siteid racegb  cesd_cut ,tcsmas, siteid racegb cesd_cut assign time assign*time
,logn,cs,Drinking risk/lognormal);
%gmix (assign siteid racegb  cesd_cut ,tcsmas, siteid racegb cesd_cut assign time assign*time
,p,cs,Drinking risk/poisson);
%gmix (assign siteid racegb  cesd_cut ,tcsmas, siteid racegb cesd_cut assign time assign*time
,gamma,cs,Drinking risk/gamma);
%gmix (assign siteid racegb  cesd_cut ,tcsmasg, siteid racegb cesd_cut assign time assign*time
,bin,cs,Drinking risk/binomial);
run;
ods rtf close;
ods listing;
quit;
run;

ods rtf;
ods listing close;

proc genmod data=four  descending;
    class pid siteid assign time (ref='1')  racegb  cesd_cut (ref='no')/param=ref   ;
    model tcsmasg = siteid racegb  cesd_cut assign time assign*time / dist=bin link=logit type3;
    repeated sub=pid / type=exch ;
           estimate "integrated" assign 1 -1/exp;
    estimate "cesd-yes" Cesd_cut 1 -1/exp;
     estimate "black" racegb 1 0 -1 /exp;
     estimate "other" racegb 0 1 -1 /exp;
      estimate "time 3" time 1 0 0 -1/exp;
      estimate "time 6" time 0 1 0 -1/exp;
      estimate "time 12" time 0 0 1  -1/exp;

     title ' genmode model /  model for TCSMASG '  ;
     title2 'PRISM study';
           run;
ods rtf close;
ods listing;
quit; run;;
```