Paper SP01-2009

Modeling the Treatment Effect on a Median of a Percent Change from Baseline in a Log-normal Variable Using SAS PROC NLMIXED

Aiming Yang, Merck & Co., Inc., Rahway, NJ
Nancy Liu, Merck & Co., Inc., Rahway, NJ
Olga M. Kuznetsova, Merck & Co., Inc., Rahway, NJ

ABSTRACT

Often, when a response percent change from baseline in a clinical parameter is not normally distributed, the medical community presents the results in terms of a median percent change and describing the treatment effect by the estimate of the difference in the treatment group medians. Such estimates are typically provided based on non-parametric analysis methods. However, if the response parameter itself is log-normally distributed, additional advantages can be derived from the parametric modeling. Following Wong et al. [1], the ANCOVA model can be fit to the log-transformed data, upon which the estimated treatment means will be back-transformed to become the estimates of the treatment group medians of the percent change from baseline. The confidence interval for the difference in medians can be provided using the delta method [1]. The introduction of SAS PROC NLMIXED [2] made this approach easily accessible as the delta method-based estimates of the confidence limits are output by the procedure, eliminating the need for tedious coding. Other capabilities of PROC NLMIXED, such as modeling the variance of the error as a function of the response, are also very useful. Finally, modeling with PROC NLMIXED can be extended to the longitudinal percent change from baseline data, thus aligning the approaches to analysis of normal and log-normal longitudinal data. These capabilities of PROC NLMIXED are illustrated in the examples (with the SAS code provided) arising from the analysis of the lipid parameters, in particular, log-normally distributed triglycerides data.

KEYWORDS

PROC NLMIXED, LOG-NORMAL DATA, DIFFERENCE IN MEDIANS, DELTA METHOD, LONGITUDINAL ANALYSIS

INTRODUCTION

For many lipid parameters measured in clinical trials, such as LDL-C, HDL-C, Total Cholesterol, non-HDL, etc., percent change from baseline is normally distributed. Because of this, these percent change variables are traditionally analyzed by the Analysis of the Variance (ANOVA) or the Analysis of the Covariance (ANCOVA) models. These models can accommodate prognostic factors and are easily executed with PROC GLM. When longitudinal data are collected, repeated measure models are fit to percent change variables using PROC MIXED. The estimates of the between-treatment differences in Least Squares (LS) Means and confidence intervals are output directly by these SAS procedures.

However, when percent change from baseline variables are not normally distributed, as is the case with triglycerides, the analytical options are less straightforward. The clinical community commonly presents the results through the treatment group medians of the percent change from baseline. The confidence intervals for the within-group medians are provided based on non-parametric methods. The treatment effect is traditionally presented through the estimate of the difference in medians and 95% Confidence Intervals (CIs) for the difference, both of which are derived non-parametrically. These non-parametric estimation methods do not incorporate important predictors, which might be desirable. Moreover, the inference is often provided by applying the ANCOVA model that incorporates the covariates to the ranked data. Thus, several different methods are applied to analyze the triglycerides data. In contrast, all the results of LDL-C analysis (within- and between – group estimates and inferences) are derived from a single ANOVA (ANCOVA) or mixed model.

To remedy the situation, advantage can be taken of the fact that the triglycerides are distributed log normally. This allows one to estimate the difference in medians of percent change from baseline that the clinical community expects to see in the parametric framework. Specifically, in a single time point analysis, the ANOVA (ANCOVA) model can be fit to the log-transformed data. The estimated treatment means on the log scale will then be back-transformed, following Wong et al. [1], to become the estimates of the treatment groups medians of the percent change from baseline. The confidence interval for the difference in medians can be provided using the delta method [1]. Although conceptually simple, the calculations the delta method involves are tedious to program in the presence of covariates and are prone to error. This explains why the method although recommended by Wong et al. [1] based on their simulations in the absence of covariates, has not been applied much in the presence of covariates.

With the introduction of SAS PROC NLMIXED, this approach became easily accessible. The parametric estimates of the medians as well as the delta method-based estimates of the confidence limits are output by the procedure, eliminating the need for tedious coding. This paper provides several examples of such applications of PROC NLMIXED.

The first example presents a typical end-of study analysis of the percent change from baseline. The flexibility of PROC NLMIXED that allows variances of the errors, varying by treatment group is illustrated in this example. Because of the non-linear transformation, it is essential to derive the treatment group LS Means weighted according to their weights in the study population before back-transforming the data, as demonstrated in this example. Example 2 illustrates other capabilities of PROC NLMIXED, such as modeling the variance of the error as a function of the response variable. In Example 3, PROC NLMIXED is used to model the longitudinal percent change from baseline data. SAS code is provided for all the examples.

EXAMPLE 1: DIFFERENT VARIANCES ACROSS THE GROUPS

Background

It has been observed in several lipid studies that the variances of some endpoints, such as triglycerides, are different across treatment groups. Heterogeneous variance can be easily incorporated using PROC NLMIXED. Below is an example to use this feature with gender as the fixed effect. Note that, in the Key Syntax, LSMeans is weighted according to the distribution of the gender in the study population for the nonlinear parameter estimation. The gender ratio of 0.4 and 0.6 is exemplified here. The model below possesses better fit statistics (AIC and BIC) compared to the one without taking the heterogeneous variance into account (results not shown); the model below also gives a more accurate estimate of the between-group treatment difference.,

Example Model:

$$\text{logTG} - \log(\text{BaseTG}) \sim \beta_1 (\text{treatment} = 1) + \beta_2 (\text{treatment} = 2) + \Upsilon (\text{gender=Male}) + \varepsilon$$

Where: logTG is the natural log of triglycerides at the end of the study;
log(BaseTG) is natural log of triglycerides at the baseline;
Treatment = 1 if on placebo, and Treatment =2 if on active tretament;
$\varepsilon \sim N(0, \sigma^2{}_i)$ i = 1,2; where σ^2 vary across treatment groups.

Key Syntax

```
proc nlmixed data=triglog cov corr    gconv=0.0000000001;
    parms m1=-0.2 m2=0.05 se2t1=0.1 se2t2=0.1 gen2=0;
    m=m1*(trt=1)+m2*(trt=2)+gender2*(gendcat=2);
    se2=se2t1*(trt=1)+se2t2*(trt=2);

    mt1=m1+gender2*0.4;
    mt2=m2+gender2*0.6;

    perc1=(exp(mt1)-1)*100;
    perc2=(exp(mt2)-1)*100;
    model diflog~normal(m,se2);
    estimate 'median of pchg in trt 1' perc1;
    estimate 'median of pchg in trt 2' perc2;
    estimate 'between trt difference in medians' perc1-perc2;
    title 'Different Errors in Different Treatment Groups';
run;
```

Example 2: ERROR DEPENDS ON THE RESPONSE

Background

Often the variances of some endpoints, such as lipoprotein (a) and some subclasses of lipids, depend on the response value. Heterogeneous variance can be easily incorporated using PROC NLMIXED. Below is an example where the variance of the error is a function of the predicted variable.

Example Model:

$$\text{logTG} - \log(\text{BaseTG}) \sim \beta_1 (\text{treatment} = 1) + \beta_2 (\text{treatment} = 2) + \Upsilon (\text{gender=Male}) + \varepsilon$$

Where: logTG is the natural log of triglycerides at the end of the study;
log(BaseTG) is natural log of triglycerides at the baseline;
Treatment = 1 if on placebo, and Treatment =2 if on active tretament;
$\varepsilon \sim N(0, var)$, where var= $pred^2 * \sigma^2$ is a function of the predicted variable.

Key Syntax

```
proc nlmixed data=lpa cov corr gconv=0.0000000001;
    parms m1=-0.2  m2=0.05 se2=0.1 gen2=0;
    m=m1*(trt=1)+m2*(trt=3)+gen2*(gendcat=2);
    mt1=m1+gen2*0.4;
    mt2=m2+gen2*0.6;
    pred = m1*(trt=1)+m2*(trt=3)+gen2*(gendcat=2);
    var = se2*pred**2;
    perc1=(exp(mt1)-1)*100;
    perc2=(exp(mt2)-1)*100;
    model diflog~normal(m,var);
    estimate 'median of pchg in trt 1' perc1;
    estimate 'median of pchg in trt 2' perc2;
    estimate 'between trt difference in medians' perc1-perc2;
    title 'Variances of Errors Depend On Responses';
run;
```

EXAMPLE 3: LONGITUDINAL DATA ANALYSIS

Background

It is very common that endpoints are measured at multiple time points during clinical studies. As a result, the longitudinal analysis model is being widely used either as the primary analysis method or as the sensitivity method in clinical studies. Below is an example of using PROC NLMIXED to model longitudinal data. PROC NLMIXED becomes very handy in this example as the parametric estimates of the medians as well as the confidence limits are output directly by the procedure, which eliminates the need for tedious coding in the presence of covariates.

Example Model:

$$\begin{aligned} logTG - \log(BaseTG) \sim\ & \alpha_{11}\,(trt=1)*(week=4) + \alpha_{12}\,(trt=1)*(week=8) + \alpha_{13}\,(trt=1)*(week=12) + \\ & \alpha_{21}\,(trt=2)*(week=4) + \alpha_{22}\,(trt=2)*(week=8) + \alpha_{23}\,(trt=2)*(week=12) + \\ & \Upsilon_1\,(week=4) + \Upsilon_2\,(week=8) + \Upsilon_3\,(week=12) + \varepsilon \end{aligned}$$

Where: Υ_1, Υ_2, and Υ_3 are the random week effects, and $(\Upsilon_1, \Upsilon_2, \Upsilon_3) \sim N(0, \Sigma)$, and Σ= 3x3 unstructured covariance matrix;
$\varepsilon \sim N(0, \sigma^2)$.

Note, since the unstructured covariance matrix is used for the random effects, we need to pre-set the variance of error (σ^2) to a very small value (0.00001). Otherwise the model will be over parameterized. This is illustrated in the Key Syntax below.

Key Syntax

```
proc nlmixed data=triglog cov corr gconv=0.0000000001;
   parms m1time1=-0.2 m1time2=-0.3  m1time3=-0.3
        m2time1=0.02 m2time2=-0.04 m2time3=0.07
        s2u1   =0.1  c12    =0.08  s2u2   =0.14
        c13    =0.06 c23    =0.06  s2u3   =0.11;
   m = m1time1*(trt=1)*(week=4)+
      m1time2*(trt=1)*(week=8)+
      m1time3*(trt=1)*(week=12)+
      m2time1*(trt=2)*(week=4)+
      m2time2*(trt=2)*(week=8)+
      m2time3*(trt=2)*(week=12)
```

```
      +u1*(week=4)+u2*(week=8)+u3*(week=12);

   p1time1=(exp(m1time1)-1)*100;
   p1time2=(exp(m1time2)-1)*100;
   p1time3=(exp(m1time3)-1)*100;

   p2time1=(exp(m3time1)-1)*100;
   p2time2=(exp(m3time2)-1)*100;
   p2time3=(exp(m3time3)-1)*100;

   model diflog~normal(m,0.00001);
   random u1 u2 u3 ~normal([0,0,0],[s2u1,c12,s2u2,c13,c23,s2u3])
   subject=an_num;
   estimate 'median of pchg in trt 1 at week 4' p1time1;
   estimate 'median of pchg in trt 1 at week 8' p1time2;
   estimate 'median of pchg in trt 1 at week 12’ p1time3;

   estimate 'median of pchg in trt 2 at week 4' p2time1;
   estimate 'median of pchg in trt 2 at week 8' p2time2;
   estimate 'median of pchg in trt 2 at week 12' p2time3;

   estimate 'between trt difference in medians at week 4' p1time1-p2time1;
   estimate 'between trt difference in medians at week 8' p1time2-p2time2;
   estimate 'between trt difference in medians at week 12' p1time3-p2time3;

   title 'Correlated Responses Within Subjects';
run;
```

CONCLUSION

SAS PROC NLMIXED provides excellent support for modeling the treatment effect on a median of a percent change from baseline in a log-normal variable in presence of covariates. Common examples of lognormal variables are some lipid parameters such as triglycerides or Lp(a). PROC NLMIXED offers rich modeling options; in particular, it allows the use of different variances of the errors for different treatment groups (Example 1). As is pointed out in the paper, because of the non-linear transformation, it is necessary to derive the treatment group LS Means weighted according to their weights in the study population before back-transforming the data.

The PROC NLMIXED outputs the estimates of the treatment group medians as well as the confidence intervals for the treatment group medians and the between-group difference in medians, sparing the statistician tedious calculations based on the delta method. Sophisticated features of the PROC NLMIXED allow modeling the variance of the error as a function of the response, which is needed, for example, in the analysis of Lp(a). PROC NLMIXED also supports longitudinal analysis of the log-normal data, as illustrated in Example 3.

Using PROC NLMIXED to estimate the difference in median in log-normally distributed variables and PROCs GLM or MIXED to estimate the difference in means for normally distributed variables, allows incorporation of the same covariates in the analysis and provides consistency of the analytic approaches across the study endpoints**.**

REFERENCES

1. Confidence Intervals for Between-Treatment Mean Difference When Analysis is on the Log Percent Scale. Proc Biopharm Sect, Am Stat Assoc. 1999;111-116. Peggy Wong, Saurabh Mukhopadhyay, Hui Quan, and Patrick Larson. Merck Research Labs RY34-3020 P.O. Box 2000, Rahway, NJ 07065

2. SAS Institute Inc. (1999), SAS OnlineDoc®, Version 9, SAS/STAT Software: Reference, Cary, N.C.: SAS Institute, Inc.

ACKNOWLEDGMENTS

The author would like to thank John Troxell, John Bowen, Sarah Shiue and Bruce Binkowitz of Merck Research Laboratories for their Advice on this paper/presentation.

CONTACT INFORMATION

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

Aiming Yang
Merck & Co., Inc.
126 Lincoln Avenue
P.O. Box 2000
Rahway, NJ 07065
Phone: 732-594-8385
e-mail: Aiming_Yang@merck.com

Nancy Liu
Merck & Co., Inc.
126 Lincoln Avenue
P.O. Box 2000
Rahway, NJ 07065
Phone: 732-594-7625
e-mail: Nancy_Liu@merck.com

Olga Kuznetsova
Merck & Co., Inc.
126 Lincoln Avenue
P.O. Box 2000
Rahway, NJ 07065
Phone: 732-594-5094
e-mail: Olga_Kuznetsova@merck.com

TRADEMARK