

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

2003 - 15th Annual Conference Proceedings

SMALL SAMPLE POWER CHARACTERISTICS OF GENERALIZED MIXED MODEL PROCEDURES FOR BINARY REPEATED MEASURES DATA USING SAS

Matthew Beckman

Walter W. Stroup

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Beckman, Matthew and Stroup, Walter W. (2003). "SMALL SAMPLE POWER CHARACTERISTICS OF GENERALIZED MIXED MODEL PROCEDURES FOR BINARY REPEATED MEASURES DATA USING SAS," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1172>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

Small Sample Power Characteristics of Generalized Mixed Model Procedures for Binary Repeated Measures Data Using SAS

By: Matthew Beckman and Walter W. Stroup
Department of Statistics, University of Nebraska, Lincoln, NE 68583-0712

Abstract

Researchers in the agricultural and biological sciences often conduct experiments with repeated measures and categorical response variables. Recent advances in statistical computing have made several options available to analyze data from these experiments. For example, SAS has several procedures based on generalized mixed model theory. These include PROC GENMOD, MIXED, NLMIXED, and the GLIMMIX macro. Inference for these procedures depends on asymptotic theory. While statistics literature contains some information about the small-sample behavior, there is much that remains unknown. This presentation will focus on Bernoulli response variables. Power characteristics are compared via simulation for several scenarios involving relatively small repeated measures experiments.

Key Words: GLMM, binary data, repeated measures, GEE, pseudo-likelihood, SAS procedures, power.

Introduction

Repeated measures data come from experiments in which measurements are observed on the same experimental unit over multiple times. Typically, these experiments involve comparisons between two or more treatments that are applied in various designs with completely randomly design and randomized block design being the most common. An example of this occurs when observing changes in animal resistance to a disease over time in response to various vaccines. In many situations, the response variable is continuous and it is reasonable to assume a normal distribution. In others, the response variable may be categorical or some other non-normal distribution.

When the data are continuous and have normal errors, then one can use a linear mixed model (LMM) such as:

$$y_{ijk} = \mu + \alpha_i + \mathbf{b}_{ij} + \gamma_k + (\alpha\gamma)_{ik} + \mathbf{e}_{ijk}$$

where y_{ijk} is the response at time k on the j th subject assigned to treatment i ; μ is the overall mean; α_i is the effect on the i th treatment; \mathbf{b}_{ij} is the random effect of the j th subject in treatment i , also known as the between subjects error; γ_k is the effect of the k th

time period; $(\alpha\gamma)_{ik}$ is the interaction effect of the i th treatment at time k ; and \mathbf{e}_{ijk} is the random effect associated with the j th subject in treatment i at time k , also known as the within subjects error. In addition, it is assumed that $\mathbf{b}_{ij} \sim \text{MVN}(0, \mathbf{I}\sigma_B^2)$, and by letting $\mathbf{e}_{ij} = [e_{ij1}, e_{ij2}, \dots, e_{ijK}]$ be the vector of within subject errors with k time periods, then $\mathbf{e}_{ij} \sim \text{MVN}(0, \Sigma)$ in which Σ is a $k \times k$ covariance matrix.

When performing the analysis of a repeated measures LMM, one can use mixed model software such as SAS[®] PROC MIXED (SAS Institute, Inc, 1999). The first step is to determine an appropriate covariance structure and estimate its variance and covariance components. Then one can assess the treatment and time effects using generalized least squares, or equivalently, by solving the mixed model equations. Littell, Stroup, and Freund (2002) described this process for PROC MIXED in detail.

In many cases, the response variable of interest is categorical – e.g. binary. In such cases, generalized linear mixed models (GLMM) are desirable, since they use the same linear combination of fixed and random effects as LMM's and inference is based on the same estimable functions one uses for LMM's, thus retaining the advantages of LMM's with respect to interpretation, but they more accurately take into account the probability distribution of the data.

Sui and Stroup (2001) presented alternative implementations of GLMM's for categorical repeated measures data available in SAS. A brief summary of the main ideas is given here. To make the transition from LMM of GLMM, one first needs to conceptualize the model in terms of the vector of random effects, \mathbf{u} and its probability distribution, the observation vector, \mathbf{y} , and the conditional distribution of \mathbf{y} given \mathbf{u} . For the LMM, $\mathbf{y}|\mathbf{u}$ is assumed $\text{MVN}(X\beta + Z\mathbf{u}, R)$ and \mathbf{u} is assumed $\text{MVN}(0, G)$. In the GLMM, the normality assumption is retained for \mathbf{u} but dropped for $\mathbf{y}|\mathbf{u}$. Instead, the quasi-likelihood of $\mathbf{y}|\mathbf{u}$ is assumed to be of the form $\frac{y\gamma(\theta, u) - b[\gamma(\theta, u)]}{\phi}$, where $\gamma(\theta, u)$ is the natural parameter, a function of $\theta = E(\mathbf{y}|\mathbf{u})$, and ϕ is a scale parameter. For binary data, the quasi likelihood has the specific form $y \log\left(\frac{\pi}{1-\pi}\right) - \log\left(\frac{1}{1-\pi}\right)$, where π is the probability of the binary outcome of interest, π is modeled as $h(X\beta + Z\mathbf{u})$, and $h(\bullet)$ the *inverse link* function. Alternatively, the model can be specified in terms of a *link* function, $g(\theta)$. Note that $h(\bullet) = g^{-1}(\bullet)$. The link function $g(\theta)$ is commonly denoted η . It represents the linear combination of fixed and random effects to be modeled directly, and is hence a function of θ to which it is reasonable to fit a linear model. The natural parameter is a typical choice. Thus, for binary data, a GLMM analogous to the repeated measures LMM given above is

$$\eta_{ijk} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \mathbf{b}_{ij} + \boldsymbol{\tau}_k + (\boldsymbol{\alpha}\boldsymbol{\tau})_{ik},$$

where $\eta_{ijk} = \log\left(\frac{\pi_{ijk}}{1 - \pi_{ijk}}\right)$, is the link function, τ_k is the effect of the k th time period, and the within-subjects error is modeled by a distribution + working correlation structure. For a more detailed discussion of GLMM's see Sui and Stroup (2001).

As Sui and Stroup discussed, SAS has four basic approaches for implementing the repeated measures GLMM described above. The specific applicability of each alternative depends on the random model effects and working correlation structure in the GLMM. The most versatile alternative is the GLIMMIX macro, which uses a pseudo-likelihood algorithm (Wolfinger and O'Connell, 1993) to augment PROC MIXED. GLIMMIX can handle GLMM analogs to any LMM that PROC MIXED can compute. For GLMM's whose covariance can be specified entirely by the working correlation structure, PROC GENMOD can be used to implement generalized estimating equations (GEE, Zeger, et al, 1998). PROC GENMOD cannot estimate variance components *per se*. For variance component models, PROC NLMIXED can be used to implement a maximum likelihood algorithm based on Gauss-Hermite Quadrature.

Which of the above options should one use? The purpose of this paper is to explore this question. To illustrate the issues, consider the simplest repeated measures model, which assumes i.i.d. within-subject errors. For LMM's, this model is equivalent to the compound symmetry models, that is, the following two SAS programs

<pre>Proc Mixed; class trt subj time; model y=trt time; random subj(trt);</pre>	<pre>Proc Mixed; class trt subj time; model y=trt time; repeated / type=cs subject=subj(trt);</pre>
---	---

yield equivalent results. For more details, see Littell, et al (2002). However, for GLMM's the equivalence of compound symmetry [*Repeated / type=cs subject=subj(trt)*] and independent errors [*random subj(trt)*] does not hold. To see the difference, consider four approaches, whose SAS code is shown in Table 1. Note that alternatives 1 and 2 represent GLMM analogs modeling between-subject error through the compound symmetry working correlation structure. Because the random effects are embedded in the working correlation and thus no random effects need to be specified, one can use PROC GENMOD's GEE option (alternative 1). One can also use GLIMMIX for this model (alternative 2). GLIMMIX uses pseudo-likelihood instead of GEE. Alternatively, one can compute the independent error, random between-subjects effect variance component model using GLIMMIX (alternative 3) or NLMIXED (alternative 4).

Table 2 shows the results on these four analyses applied to an example data set with two treatments, twenty subjects per treatment, each observed at five time periods. Note that each procedure produces a unique result. For the compound symmetry model, GENMOD (GEE) and GLIMMIX (pseudo-likelihood) produce similar (but not identical)

estimates of ρ and the p -value for treatment-by-time interaction. For the variance components model, GLIMMIX and NLMIXED (Gauss-Hermite Quadrature) produce moderately different between-subject variance estimates ($\hat{\sigma}_b^2$) and very different treatment-by-time p -values.

With all these options available and conflicting results, one wonders which approach is right. To address this question, we need to establish what are the criteria for determining a “right” approach? This paper will use the following criteria:

1. Control over type I error. A “good” approach, performing at the nominal rejection rate, should reject the null hypothesis of interest with probability equal to the stated α -level.
2. Given the approaches that effectively control type I error, i.e. reject at the nominal rejection rate when the null hypothesis is true, we want the method that maximizes power when treatments are different.

To compare the four GLMM approaches for binary data, a simulation study was performed. Simulation is required for several reasons. First, all four approaches are iterative procedures that lack closed form solutions. Second, the test statistics for all four procedures depend on asymptotic theory whose small-sample behavior is not well documented. The focus of this paper is to study the small-sample behavior of these procedures under conditions typical of their likely use in agricultural research.

Simulation Study Materials and Methods

The models and procedures described in Table 1 were evaluated using simulated data from repeated measures designs with two treatments, a completely randomized between-subjects design, a binary response, and five times of measurement per subject. The structure of the simulated data was suggested by companion animal vaccine trials and the need to assess prospective methods of analysis. The simulated data sets used varying numbers of animals (subjects) per treatment (10, 20, and 40 animals/trt), different patterns of treatment differences (shown below) and within-subject autocorrelation ($\rho = 0$ or 0.75). Only results for data sets with 20 and 40 animals/trt are discussed. With 10 animals/trt, GENMOD (GEE), GLIMMIX, and NLMIXED all performed poorly, showing high (over 50%) rates of non-convergence and other symptoms of unreliable computations.

Data were generated as follows. The SAS RANNOR function was used to generate between subjects errors, \mathbf{b}_{ij} and within subjects errors, \mathbf{e}_{ijk} . The \mathbf{b}_{ij} generated were i.i.d. $N(0, \sigma_b^2)$ with σ_b^2 set to 5. The \mathbf{e}_{ijk} were generated according to an AR(1) process within each subject. If $\rho = 0$ then the \mathbf{e}_{ijk} were i.i.d. $N(0,1)$. If $\rho = 0.75$ then the \mathbf{e}_{ijk} were generated as $e_{ijk} = \rho e_{ij,k-1} + w_{ijk}$ where w_{ijk} were i.i.d. $N(0,1)$. These values were used to calculate $\eta_{ijk} = \mu_{ik} + b_{ij} + e_{ijk}$, where μ_{ik} is determined by the pattern of treatment differences described below. The normally distributed η_{ijk} were converted to

probabilities using the logit inverse link, $\pi_{ijk} = \frac{e^{\eta_{ijk}}}{1 + e^{\eta_{ijk}}}$. Using the probability π_{ijk} the binary observations y_{ijk} were generated using the SAS RANBIN function.

Note that $E(\eta_{ijk}) = \mu_{ik}$ and $E(\pi_{ijk}) = \pi_{ik}$. Also, μ_{ik} and π_{ik} are functionally related by the logit link, i.e. $\mu_{ik} = \log\left(\frac{\pi_{ik}}{1 - \pi_{ik}}\right)$. Simulated data were generated using the following four patterns of simulated differences:

$$\pi_1 = \begin{bmatrix} 0.1 & 0.1 \\ 0.1 & 0.1 \\ 0.1 & 0.1 \\ 0.1 & 0.1 \\ 0.1 & 0.1 \end{bmatrix} \quad \pi_2 = \begin{bmatrix} 0.1 & 0.1 \\ 0.1 & 0.2 \\ 0.1 & 0.3 \\ 0.1 & 0.3 \\ 0.1 & 0.3 \end{bmatrix} \quad \pi_3 = \begin{bmatrix} 0.1 & 0.1 \\ 0.1 & 0.3 \\ 0.1 & 0.5 \\ 0.1 & 0.5 \\ 0.1 & 0.5 \end{bmatrix} \quad \pi_4 = \begin{bmatrix} 0.1 & 0.1 \\ 0.1 & 0.45 \\ 0.1 & 0.8 \\ 0.1 & 0.8 \\ 0.1 & 0.8 \end{bmatrix}$$

Each π_w ($w = 1, 2, 3, 4$) contains the π_{ik} 's which represent the probability of a favorable outcome for the i th treatment of the k th time. Thus, each column represents a treatment and each row a time.

The first set, π_1 , represents the case of no treatment difference, no time effect, and no TRTxTIME interaction. It was used to assess control over type I error. The sets π_2 , π_3 , and π_4 were used to assess power.

For convenience in characterizing power as a function of increasing size of treatment difference, we calculated the following statistic, based on the non-centrality parameter, denoted as ϕ_w :

$$\phi_w = \sum_{j=1}^5 (\pi_{1j} - \pi_{2j})^2, \text{ where } w \text{ corresponds to the respective set.}$$

Thus the resulting ϕ_w 's are as follows:

$$\phi_1 = 0.0 \quad \phi_2 = 0.13 \quad \phi_3 = 0.52 \quad \phi_4 = 1.59$$

In evaluating type I error, i.e. using set π_1 , several options were considered for each procedure. The GEE procedure using PROC GENMOD was computed for both the

exchangeable and unstructured working correlation structures. Only results for the exchangeable structure appear because with the unstructured model, the procedure failed to converge for a large number of the simulated data sets. For the GLIMMIX macro, we examined both the compound symmetry and unstructured working correlation models. We also compared the power characteristics of the default, model-based test statistics to 1) the Kenwood-Roger degrees of freedom adjustment, and 2) the “sandwich” estimator (SAS PROC MIXED “empirical” option, see Diggle, et al., 1994). Finally, we analyzed the data by using the LMM to analyze the binary responses directly using PROC MIXED, to see how it performs relative to the other methods.

Because 500 simulated experiments were generated for data set π_1 , one would expect an observed rejection rate within ± 0.02 , or between 3% to 7% rejections, if the test is actually performing at a nominal $\alpha=0.05$ level. A rejection rate less than 0.03 suggests an excessively conservative test, whereas a rate exceeding 0.07 suggests inadequate type I error control. For GENMOD and GLIMMIX, some procedures yielded observed type I error rates greater than 0.07. These options were dropped from subsequent evaluation of power, as a statistically significant result would only be credible if it comes from a procedure that adequately controls Type I error.

Results and Discussion

Table 3 shows the observed rejection rates for H_0 : no TRT \times TIME interaction at $\alpha = 0.05$, using the pattern of treatment difference set π_1 and $\rho = 0$. For set π_1 , this provided an indication of type I error control. With 20 subjects per treatment, the NLMIXED procedure yielded an observed rejection rate of 0.02, below the 0.03-0.07 range one would expect for $\alpha=0.05$, suggesting an excessively conservative test. Two procedures, GLIMMIX with random between subjects effects and GLIMMIX using the sandwich estimator (empirical option), yielded unacceptably high observed type I error rates. With 40 subjects per treatment the observed rejection rate of NLMIXED improved to 0.048. However, increasing the number of subject per treatment did not improve either the GLIMMIX variance component (random within subject effect) model or empirical option: their rejection rates remained inflated. The type I error rate inflation for GLIMMIX variance component model was particularly severe: above 0.17 in both cases. This is consistent with other literature (e.g. Breslow and Clayton, 1993) suggesting that pseudo-likelihood based variance component estimates for GLMM's with binary data can be severely biased. Because the test statistics depend on the variance component estimates, the result here is severely upwardly biased F-statistics. For this reason, the GLIMMIX variance component approach was deemed unsuitable for use with binary data, and was dropped from the power analysis.

Power, i.e. rejection rate when H_0 is false, was monitored for increasing differences represented by π_2 , π_3 , and π_4 . For convenience, power was monitored as a function of ϕ_w as defined above. Methods of analysis that failed to control type I error, i.e. with observed rejection rates exceeding 0.07 under π_1 (H_0 true) were not included in the power analysis.

We assessed the rejection rates and the power characteristics for the TRTxTime interactions, the TRTxLinear effects, and the TRTxQuadratic effects by creating Power curves for the various methods. These graphs can be seen in Figures 1 through 6.

Figures 1 through 3 show power as a function of ϕ_w for the over test of TRTxTime interaction and for the TRTxLinear effect of time and TRTxQuadratic contrasts, respectively, for 20 subjects per treatment. For this case, applying LMM directly to the binary observations via PROC MIXED yielded the greatest power. The GEE method (PROC GENMOD) yielded the worst power, with pseudo-likelihood (CS, GLIMMIX) in the middle. Figures 4 through 6 show the results for 40 subjects per treatment. The only major difference between 20 and 40 subjects is that the power differences between GEE and pseudo-likelihood appear to be attenuated somewhat for the 40 subject/trt case, but both are still less powerful than direct analysis using PROC MIXED.

Figures 1 through 6 assume independent errors, i.e. the autocorrelation $\rho = 0$. However, in a “live” analysis one cannot assume $\rho = 0$. Therefore, a second power simulation was done with $\rho = 0.75$, a relatively large autocorrelation.

Again, we assessed the rejection rates and the power characteristics for the TRTxTime interactions, the TRTxLinear and the TRTxQuadratic contrasts. Figures 7 and eight show the results for the TRTxTime test for 20 and 40 subjects respectively. The TRTxLinear and TRTxQuadratic results follow the same pattern as the $\rho = 0$ case and are therefore not shown here. As for the $\rho = 0$ case, PROC MIXED yields maximum power, GEE yields minimum power. Pseudo-likelihood is in the middle. The difference between pseudo-likelihood and GEE appears somewhat attenuated as the number of subjects per treatment increases.

Conclusion

The primary purpose of this study was to gain some understanding of the small sample behavior of alternative methods to analyze binary data from small repeated measures experiments. From the study, several conclusions can be drawn. These are:

- The SAS-available GLMM algorithms considered in this paper performed poorly with fewer than 20 subjects per treatment. They have high rates of non-convergence and, in any event, the power characteristics are poor even for gross treatment differences. This raises significant questions about the viability of studies with few subjects and binary data.
- GLMM variance component models for binary data are poorly estimated by pseudo-likelihood methods, e.g. as implemented by the SAS GLIMMIX macro. Biased variance component estimates observed in this study are consistent with other GLMM literature. These cases also yielded poor type I error control. Gaussian quadrature, e.g. as implemented by SAS PROC NLMIXED performs acceptably with variance component GLMM's provided the sample size is adequate: tests appeared to be conservative with 20 subjects per treatment, but reasonable with 40 subjects per treatment.

- GLMM's with compound symmetry working correlation structures effectively control type I error, using either pseudo-likelihood (GLIMMIX) or GEE (e.g. via PROC GENMOD). Pseudo-likelihood methods yield greater power than GEE for tests of treatment by time interaction.
- Using an unstructured working correlation matrix rather than compound symmetry had little consistent impact on Type I error control and power for the pseudo-likelihood procedure. This is true regardless of whether the underlying data are auto correlated or not. On the other hand, the GEE procedure does not work well with an unstructured working correlation matrix for binary data: convergence rates were unacceptably low.
- In all cases, fitting an LMM directly to the binary response, e.g. via PROC MIXED yielded superior power as well as acceptable type I error control.

This last result warrants additional comment. The LMM using PROC MIXED fits a different model than the logistic GLMM fitted by the other procedures. In this sense, one can argue that the comparison is not exactly fair. However, from a practical point of view, those making decisions regarding the presence or absence of a treatment effect on changes in the probability of a favorable outcome over time will act on conclusions from the LMM and the logistic GLMM interchangeably. In this sense, the LMM approach appears to be more powerful in revealing treatment effects without compromising type I error control.

In addition, this simulation study has raised several more questions. First regards the performance of PROC MIXED. Because these simulations used an underlying normal process to generate binary observations, it is possible that this gave the LMM a comparative advantage. We tried alternative simulated data sets, not shown here, and the PROC MIXED LMM retained its advantage. However, the simulation is clearly not exhaustive, and there may be a point at which PROC MIXED's advantage disappears. This warrants further investigation.

The second topic for additional investigation concerns auto correlated errors. In principle, AR(1) + variance component GLMM's similar to those LMM's discussed in Littell, et. al. (2002) can be analyzed. The pseudo-likelihood GLMMIX approach was observed and did as poorly in this study as the variance component only model. However, PROC NLMIXED could be used to analyze the model using Gaussian quadrature. We developed a program, but it took far too long to run to be able to include it in this simulation study. In addition, this study focused only on methods readily available in SAS. We did not consider Bayesian methods, e.g. those using MCMC procedures, because the software is considerably less developed and accessible to the average statistical consultant. However, these methods may have real advantages that warrant further study. As statistical computing improves, these methods will no doubt be much more accessible.

Finally, these results apply to binary data. To what extent do these results extend to multinomial data? While GLIMMIX is restricted to binary GLMM's, PROC GENMOD can compute multinomial GEE's and PROC NLMIXED can compute cumulative logit and cumulative probit models. Little is known about how the small

sample behavior of these procedures compare. Clearly, there is much about GLMM's for repeated measures and categorical data we still do not know.

References

- Breslow, N.E. and D.G. Clayton. 1993. "Approximate inference in generalized linear mixed models." *J Amer. Statist. Assoc.* **88**: 9-25
- Littell, R.C., Walter W. Stroup, and Rudolf J. Freund. 2002. *SAS For Linear Models*. Cary, NC: SAS Institute
- SAS Institute, Inc. 1999. *SAS Online Doc, Version 8*. Cary, NC: SAS Institute.
- Sui, Yaobing and W.W. Stroup. 2001. "A Comparison of Some Methods to Analyze Repeated Measures Ordinal Categorical Data." *Proceedings: 13th Annual Conference on Applied Statistics in Agriculture*. Manhattan, KS: Kansas State University Statistics Department.
- Wolfinger, R.D. and M. O'Connell. 1993. "Generalized linear mixed models: a pseudolikelihood approach." *J. Statist. Comput. Simul.* **48**: 233-243.
- Zeger, SC., K.-Y. Liang, and P.S. Albert. 1988. "Models for longitudinal data: a generalized estimating equation approach." *Biometrics* **44**: 1049-1060.

Tables and Graphs

Table 1: SAS Program Code for Various Analyses

1. GEE, CS working correlation matrix

```
Proc genmod data=binary_example;
  by expt;
  class trt time subj;
  model y = trt|time / dist = binomial link=logit;
  repeated subject = subj(trt)/ type=exch;
```

2. Pseudo-likelihood, CS working correlation matrix

```
%Glimmix (data= binary_example, procopt=method=reml,
  stmts=%str (
  class subj trt time;
  model y = trt|time;
  repeated / type=cs subject=subj(trt);
```

3. Pseudo-likelihood, random between-subject effect

```
%Glimmix (data= binary_example, procopt=method=reml,
  stmts=%str (
  class subj trt time;
  model y = trt|time;
  random subj(trt);
```

4. Gauss-Hermite Quadrature

```
Proc nlmixed qpoints=21 data=binary_example gconv=0.0000000001;
  by expt;
  parms b0=0 a1=0 b1=0 b2=0 b3=0 b4=0 ab11=0 ab12=0 ab13=0 ab14=0
  s2b=1;
  eta = b0 + a1*t1 + b1*w1 + b2*w2 + b3*w3 + b4*w4 + ab11*t1*w1 +
  ab12*t1*w2 + ab13*t1*w3 + ab14*t1*w4 + bsel;
  p = exp(eta)/(1 + exp(eta));
  model y~binomial(1,p);
  random bsel ~normal([0],[s2b])
  subject = animal;
```

Table 2: Summary of Analyses of Two-Treatment Binary Example Using Table 1 Programs

Type of Analysis	$\hat{\rho}_{cs}$	$\hat{\sigma}_b^2$	$F_{TRTxTIME}$	$\chi^2_{TRTxTIME}$	$p > F$
GEE	0.2511	NA	N.A.	8.23	0.0835
GLIMMIX- CS	0.2577	NA	2.24	N.A.	0.0675
GLIMMIX-Rand BSE	NA	2.4453	3.43	N.A.	0.0103
NLMIXED-Rand BSE	NA	2.6926	2.21	N.A.	0.0700

Table 3. Observed Type I Error (Rejection) Rates of Procedures using Simulated Data Set π_1

20 subject/trt	Rejection Rate	40 subjects/trt	Rejection Rate
GENMOD-CS GEE*	0.038	GENMOD-CS GEE*	0.060
GLIMMIX-CS	0.042	GLIMMIX-CS	0.066
GLIMMIX -random betw-subj	0.174	GLIMMIX -random betw-subj	0.178
NLMIXED -random betw-subj	0.020	NLMIXED -random betw-subj	0.048
GLIMMIX-CS ddfm=kr	0.040	GLIMMIX-CS ddfm=kr	0.027
GLIMMIX-UN	0.033	GLIMMIX-UN	0.050
GLIMMIX-UN ddfm=kr	0.043	GLIMMIX-UN ddfm=kr	0.070
GLIMMIX-UN empirical	0.083	GLIMMIX-UN empirical	0.077
MIXED	0.053	MIXED	0.047

* 2/500 failed to converge

Figure 1. Power Curves for Various Methods, test for Treatment \times Time interaction, 20 subjects per treatment, within-subject correlation $\rho=0$

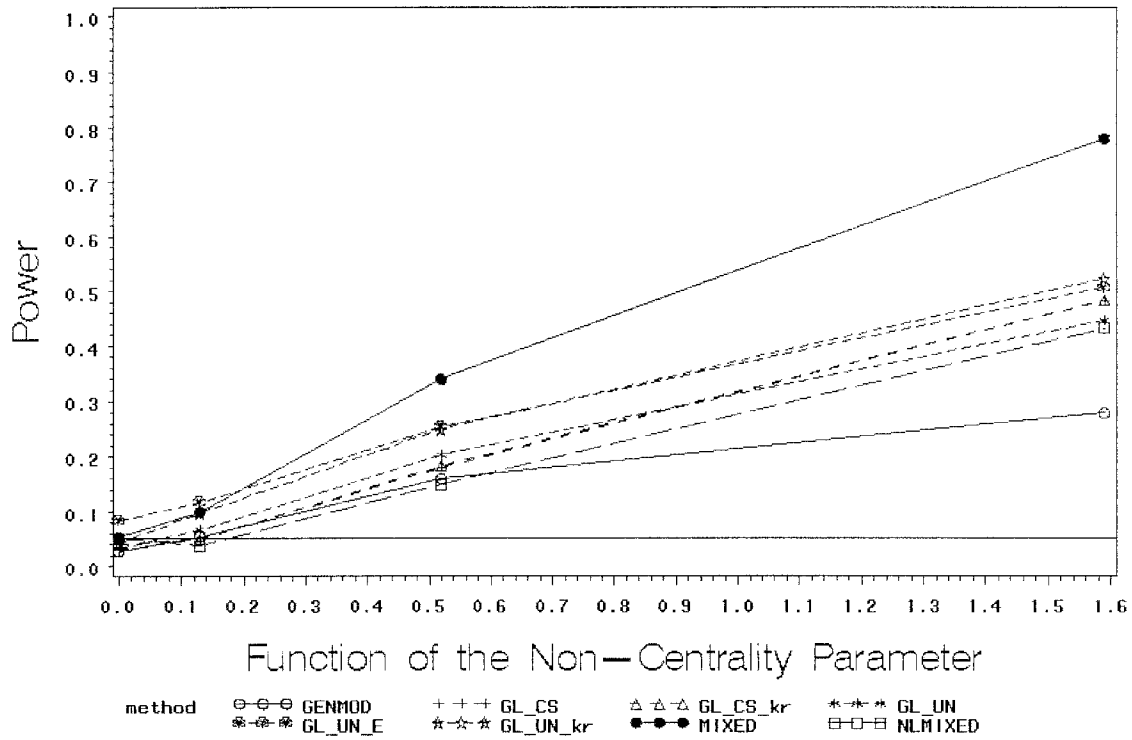


Figure 2. Power Curves for Various Methods, test for Treatment \times linear time interaction, 20 subjects per treatment, within-subject correlation $\rho=0$

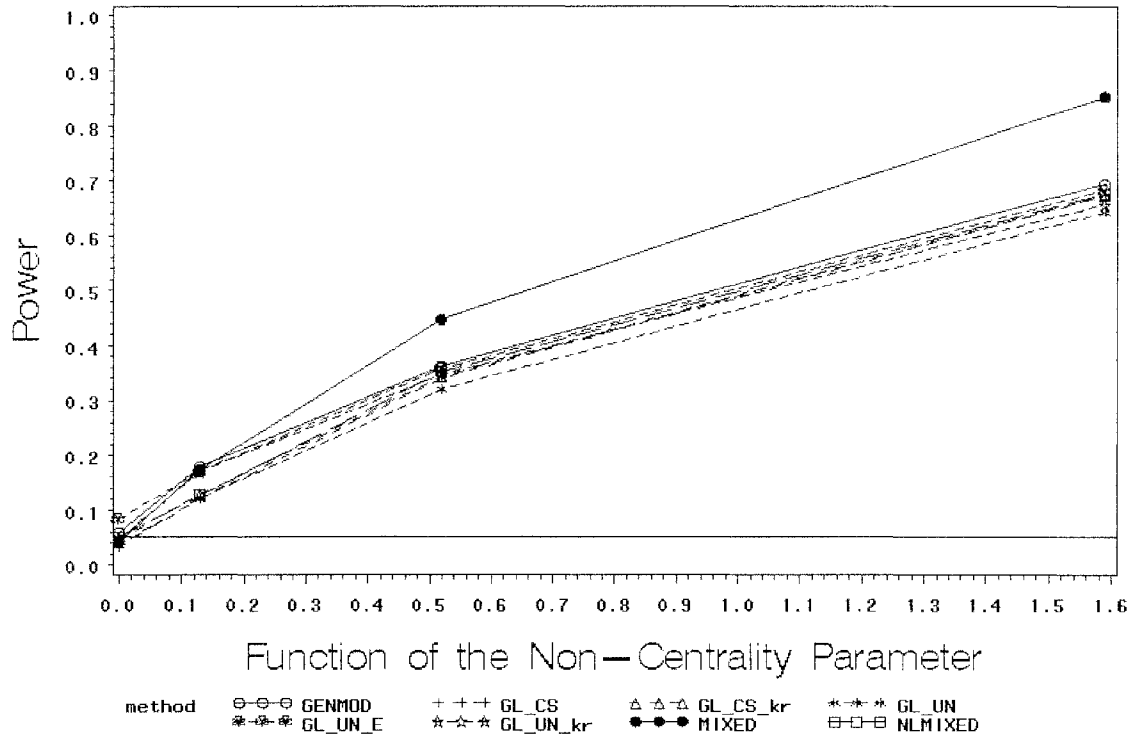


Figure 3. Power Curves for Various Methods, test for Treatment \times Quadratic time interaction, 20 subjects per treatment, within-subject correlation $\rho=0$

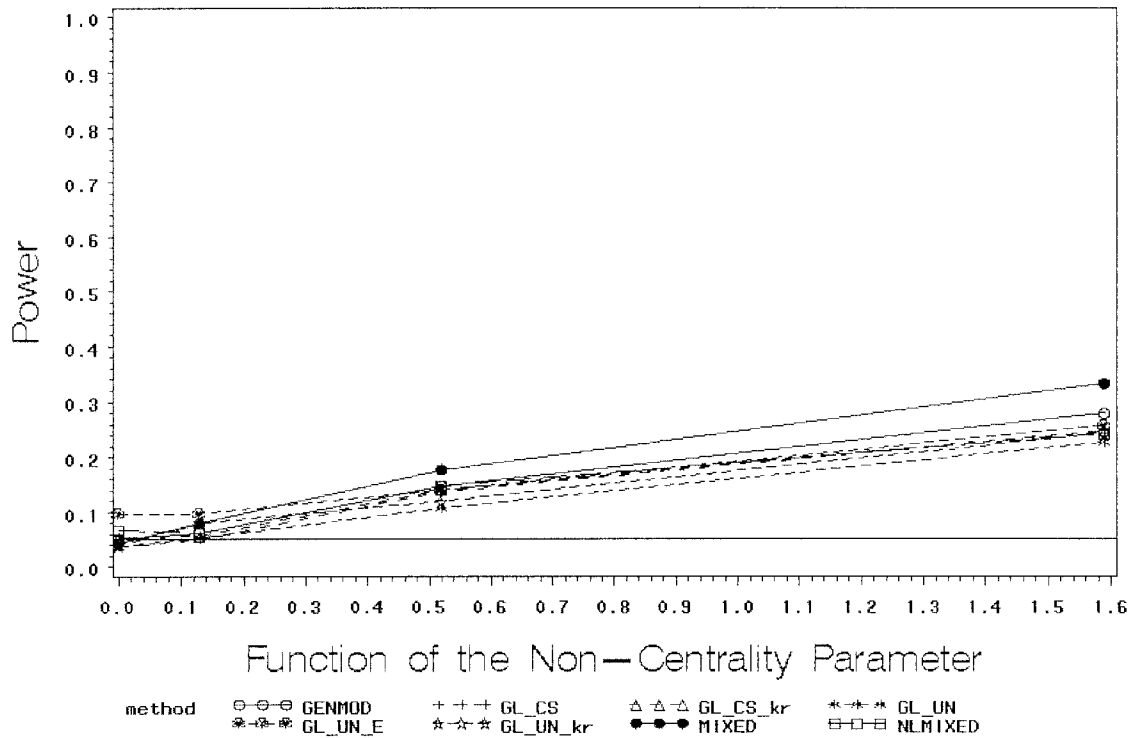


Figure 4. Power Curves for Various Methods, test for Treatment \times Time interaction, 40 subjects per treatment, within-subject correlation $\rho=0$

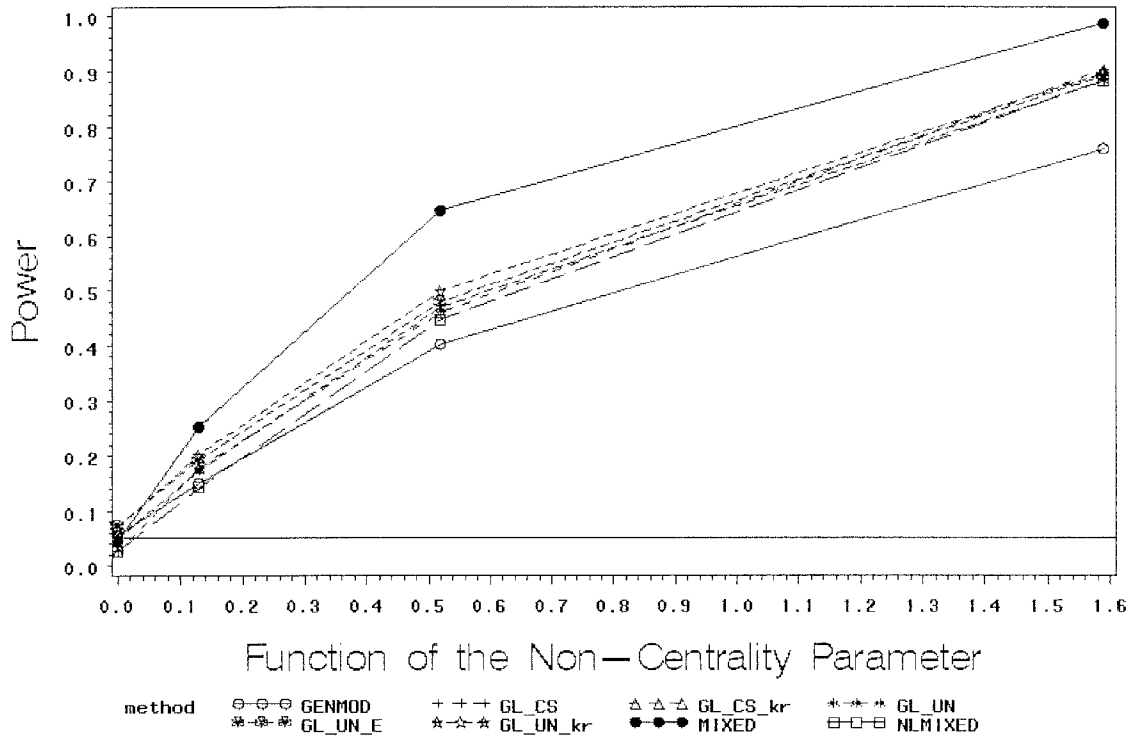


Figure 5. Power Curves for Various Methods, test for Treatment \times linear time interaction, 40 subjects per treatment, within-subject correlation $\rho=0$

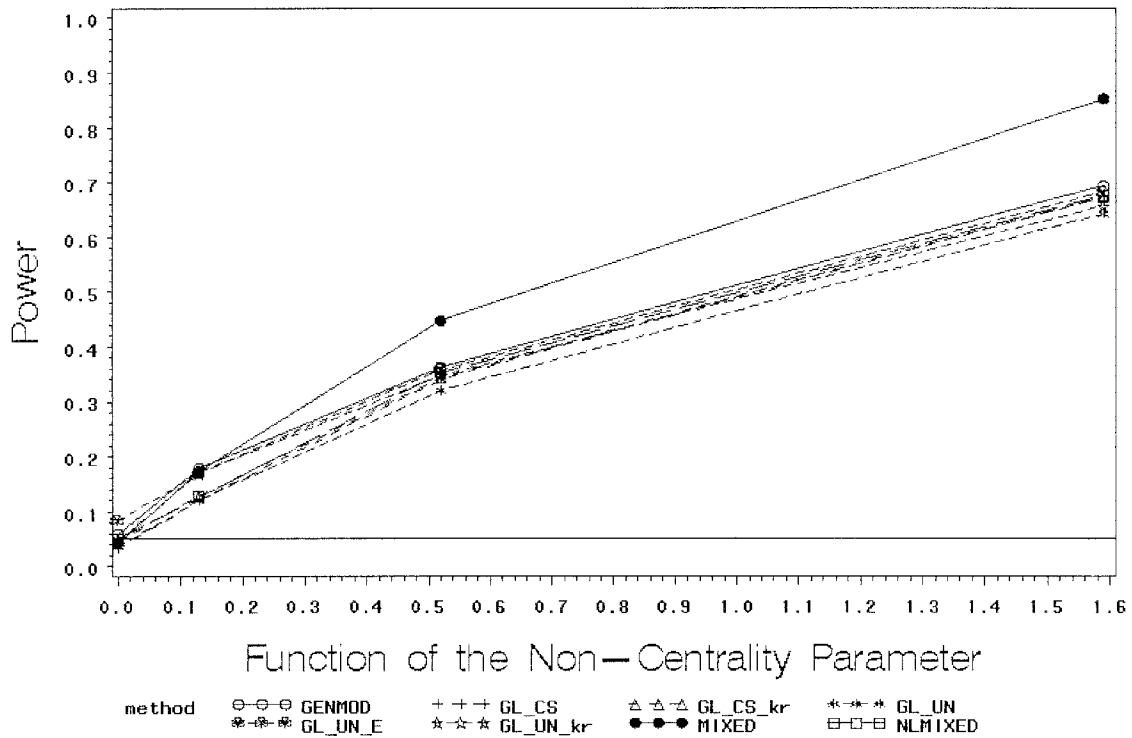


Figure 6. Power Curves for Various Methods, test for Treatment \times Quadratic time interaction, 40 subjects per treatment, within-subject correlation $\rho=0$

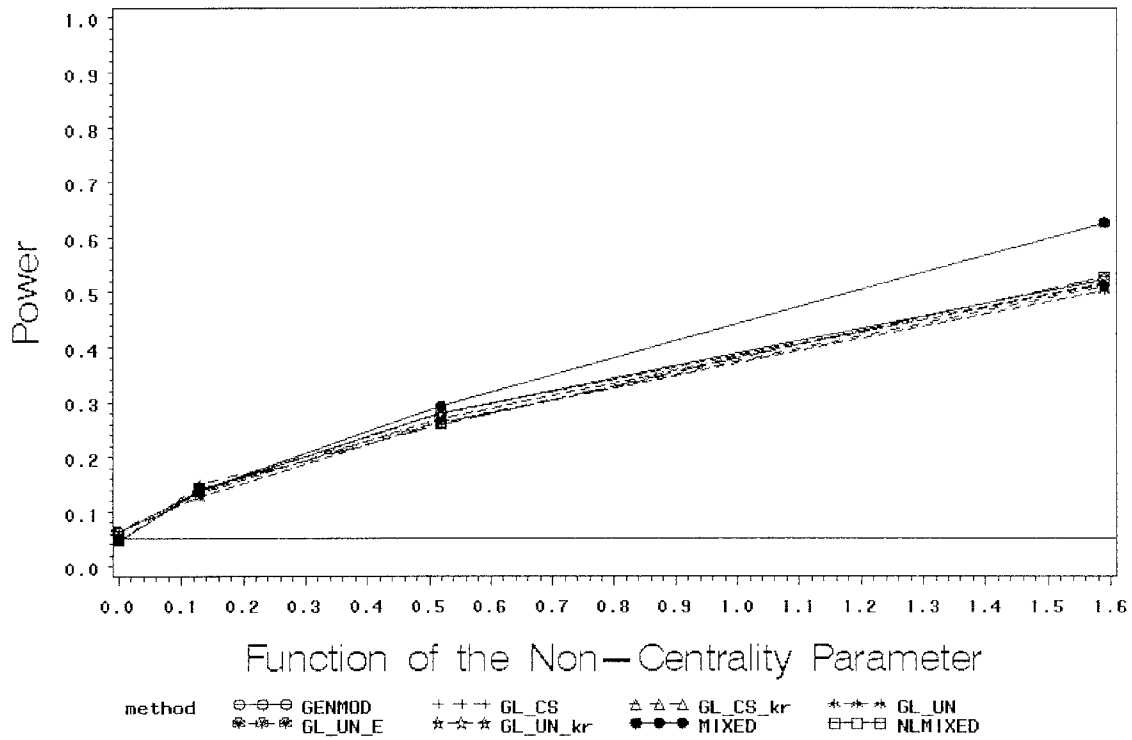


Figure 7. Power Curves for Various Methods, test for Treatment \times Time interaction, 20 subjects per treatment, within-subject correlation $\rho=0.75$

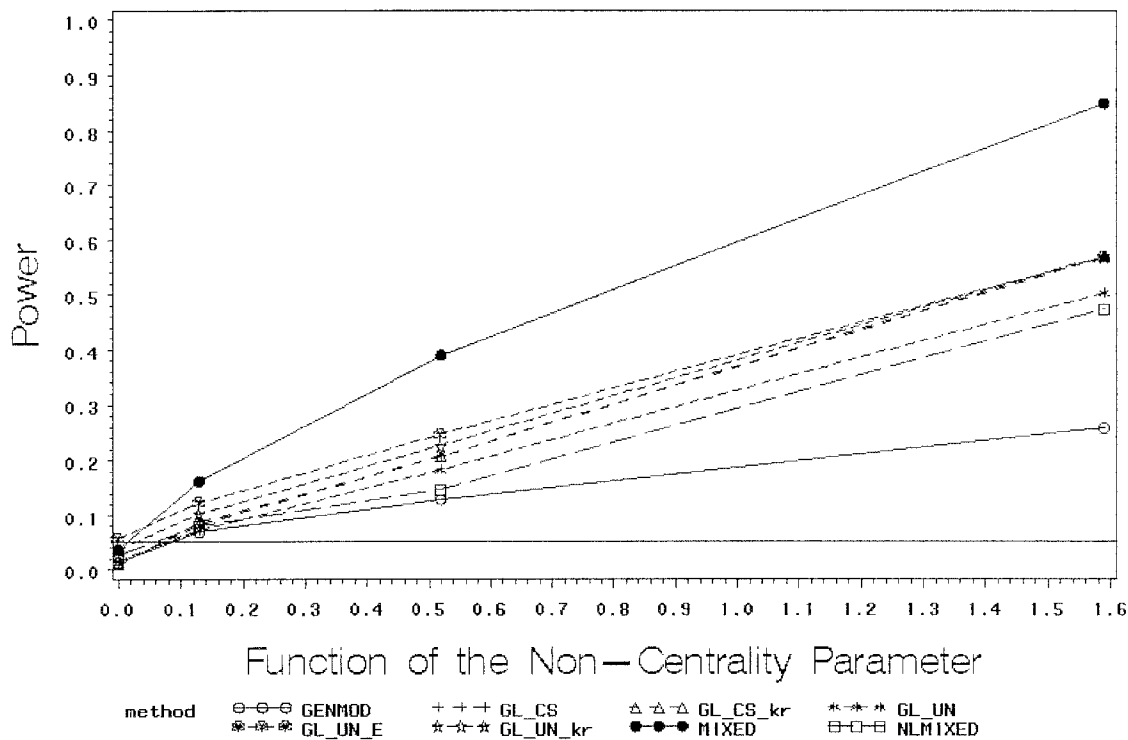


Figure 8. Power Curves for Various Methods, test for Treatment \times Time interaction, 40 subjects per treatment, within-subject correlation $\rho=0.75$

