LONGITUDINAL ANALYSIS OF RENAL FUNCTION USING ZIP GEE ON OLT TRANSPLANT PATIENTS UNDERGOING NAC PROPHYLAXIS

by

Shekhar Mehta

B.S. Statistics, The George Washington University, 2000

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2006

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

Shekhar Mehta

It was defended on

July 31, 2006

and approved by

Thesis Advisor: Abdus Wahed, PhD
Assistant Professor
Department of Biostatistics
Graduate School of Public Health, University of Pittsburgh

Committee Member: Nancy Sussman, PhD
Assistant Professor
Department of Environmental and Occupational Health
Graduate School of Public Health, University of Pittsburgh

Committee Member: Ibtesam Hilmi, MD
Assistant Professor
Department of Anesthesiology
School of Medicine, University of Pittsburgh

Committee Member: Jong-Hyeon Jeong, PhD
Assistant Professor
Department of Biostatistics
Graduate School of Public Health, University of Pittsburgh

LONGITUDINAL ANALYSIS OF RENAL FUNCTION USING ZIP GEE ON OLT TRANSPLANT PATIENTS UNDERGOING NAC PROPHYLAXIS

Shekhar Mehta, M.S.

University of Pittsburgh, 2006

Organ dysfunction is associated with oxidative stress following Orthotropic Liver Transplant (OLT) surgery. N-Acetyl Cysteine (NAC) is an acetylated form of the amino acid cysteine. NAC is known to replenish glutathione in the bloodstream which helps relieve cell damage caused by oxidative stress. NAC was used in a placebo controlled study to discover its effects on organ dysfunction caused by oxidative stress following OLT surgery. A standard NAC treatment, as used to treat acetaminophen toxicity, was used as a treatment during surgery. Measures of hepatic and renal dysfunction were recorded at unequally spaced time-points for a follow-up time of one year. The Generalized Estimating Equation (GEE) approach was used to model continuous hepatic responses. Discrete renal dysfunction responses are shown to follow a unique distribution. This unique distribution was accommodated by the GEE procedure proposed by Liang and Zeger to produce consistent and efficient estimates of the treatment effect of NAC. The estimates produced contradictory results for a hypothesized protective effect of NAC against hepatic dysfunction. The public health relevance of this work is that NAC treatment, if shown to be efficacious with respect to renal function, can benefit over six thousand OLT patients each year.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Toxic radicals are known to have a detrimental effect on internal organ tissue[5]. This cell damage is especially relevant for organ transplant patients as their immune systems are weakened to help prevent organ rejection. Oxidative stress has been shown to be related to renal dysfunction in early post-operative stages of Orthotropic Liver Transplant (OLT)[5]. N-Acetyl Cysteine (NAC) is the acetylated form of the amino acid cysteine. It is rich in thiol groups. Thiol groups of glutathione act as competitive inhibitors of toxic radicals for protein substrates. NAC helps replenish the thiol groups of glutathione [2]. NAC also effects the concentration of nitric oxide in the blood. Nitric oxide is a powerful vasodilator. Vasodilatation is important following liver transplant because it increases dieresis, which helps clear the blood of toxic chemicals. Decreases in urinary output are associated with renal dysfunction [14]. Renal dysfunction is associated with post-operative mortality rates of OLT patients. The one month mortality rate for patients who develop renal dysfunction is fifty percent, while the mortality rate for those with normal renal function is close to thirty percent [4].

In analysis of the treatment NAC on renal and hepatic functioning status the patient population consists of adults who have undergone OLT transplant surgery at the University of Pittsburgh Medical Center. A total of 82 patients were recruited. Of the 82 patients 8 were dropped from the analysis due to cancellation of the procedure, death within the first three days following surgery, or unknown treatment status. There are several patient characteristics that may influence organ dysfunction. These patient characteristics include an indicator for treatment status, age in years, gender (designated by an indicator variable for male gender), amount of red blood cells transfused during surgery, and body mass index. The organ donor characteristics of age, in years, and serum sodium level of the donor liver were also recorded. Tatsuka et. al. have proven a strong correlation between donor serum sodium levels and liver graft performance. [13]

This donor effect is intuitively expected. The status of the donor liver will greatly affect the condition of the transplant patient. Patients are randomized at a 1 to 1 ratio to either the treatment or placebo group. In the analyzed patient population there are 39 assigned to placebo, and 35 assigned to the NAC treatment. The NAC was prepared by the UPMC Presbyterian Pharmacy. The initial treatment consisted of a load dose of 140mg/kg NAC, given by IV, over one hour prior to surgery. Then 70mg/kg was administered over one hour every 4 hours for 12 additional doses. This dosage regime is the standard used for acetaminophen induced toxicity at UPMC medical center. The patients in the placebo groups received an identical administration of normal saline solution. All patients were followed for one year with the goal of ascertaining the effect of NAC on organ dysfunction.

The hepatic response data consists of repeated measurements of Alanine Aminotranferase (ALT), Aspartate Aminotransferase (AST), Bilirubin, and International Normalized Ratio (INR). The time intervals for hepatic responses are unequal. A baseline measurement was recorded before the surgery (day 0). Measurements were then taken at the following time-points: each post-operative day for the first week (day1 - day7), at 1month (day 30), 3month (day 90), 6month (day 180), and 1year (day 360) for a total of twelve time-points. ALT is an enzyme that is released into the bloodstream by the liver after hepatocytes are damaged or destroyed. ALT catalyzes the transfer of an amino group from alanine to alpha-ketoglutarate, and acts as a nitrogen buffer. This transfer of amino groups is important to renal functioning as well since kidney damage causes the retention of toxic, nitrogen rich urea. AST is also a liver enzyme released when hepatocytes are damaged. AST is found in cardiac and other organ tissues. Males often have higher levels of AST as it is found in a range of muscle tissue types. Bilirubin is a product of the destruction of red blood cells. It is removed from the bloodstream by the liver. Bilirubin measures in the blood are indicative of liver functioning status. INR is a standardized measure of prothrombin time. Prothrombin time measures blood coagulation time. The standardization of INR accounts for differences in the clinical lab conducting the tests. Each lab uses a variably prepared tissue factor called thromboplastin, which is a protein used to initiate blood coagulation in the test.

$$INR = \left(\frac{PT_{test}}{PT_{normal}} \right)^{ISI}$$

International Sensitivity Index (ISI) is determined by the manufacturer of the tissue factor. The INR of the patient population ranges from 0.6 to 3.7 with normal values between 0.8 and 1.2. . These four measurements together give a sufficient indication of liver functioning status.

The renal response consists of repeated measure of post-operative Acute Renal Failure Scores (ARF) as designated by the RIFLE criteria [7]. The ARF scores are based on levels of urinary output and serum creatinine. There are five levels of ARF score: 0 no risk; 1 risk ; 2 Injury ; 3 Failure ; 4 Persistent ARF ; and 5 End Stage Renal Disease (ESRD). The ARF scores were also recorded at unequal time intervals. A baseline measurement was taken before the OLT surgery at day0. ARF responses were then recorded at the following times: post-op at the end of treatment day 1, day 7, day 14 , day 21, 1month (day 30), 3month (day 90), 6month (day 180), and 1 year (day 360). The objective of this thesis is to analyze the effect of NAC on hepatic and renal functioning status of OLT transplant patients using the GEE approach. Preliminary graphical analysis of the responses will be conducted to try to understand the trend of responses over the study time. The patient characteristics between treatment groups will be compared to ascertain randomness. The procedure of GEE will be implemented to estimate the main treatment effect on both the hepatic and renal responses. The effect of the treatment if any will give evidence toward a protective effect of NAC on OLT transplant patients

1.1 THE GENERALIZED ESTIMATING EQUATION APPROACH

The most widely known method to model repeated responses that are assumed to depend on a mixture of both discrete and continuous covariates is an extension of the group of Generalized Linear Models proposed by Nelder and Wedderburn[12]. This extension was originally proposed by the seminal paper *Longitudinal Analysis using Generalized Linear Models* by Liang and Zeger [10]. Liang and Zeger adapted the Generalized Linear Model framework to incorporate longitudinal response data. These extensions use a working correlation structure that deals with the correlated longitudinal responses to consistently model the parameter values. Main effects estimators are assumed to be consistent without regard to the assumed correlation structure [1, p.471]. For example, suppose there is a collection of random variables Y_{ij} representing the value of a response for the i^{th} patient at replication j = 1 to *p*. Consider the normally distributed case where $Y_i \sim N\left(\mu_i, \Sigma_i\right)$. Where μ_i = E(Y_{ij}), a 1 x *p* vector of repeated responses and Var(Y_{ij}) = Σ a *p* x *p* matrix with the marginal variances on the main diagonal and covariances among the replications elsewhere. Assuming the response depends upon a vector of length *m* of covariate patient characteristics. The model is written as follows:

$$g\left(E\left[Y_{ij}\right]\right) = \sum_{k=1}^{m} \beta_{ik} X_{ik}$$

for a suitable link function g(-). For the normal case the link function is the identity link. The response is directly related to the linear function of covariates. Estimates of the coefficients of the covariates are given by solving the following score equation [1]:

$$S\left(\hat{\beta}\right) = \sum_{i=1}^{N} D_i' \, V_i^{-1} \left(y_i - \mu\right) = 0$$

$$V_i = A_i^{\frac{1}{2}} R_i\left(\hat{\alpha}\right) A_i^{\frac{1}{2}}$$

And the variance of the estimate is given by:

$$V\left(\hat{\beta}\right) = M_0^{-1} \, M_1 \, M_0^{-1}$$

$$\textit{where } M_0 = \sum_{i=1}^{N} D_i' \, V_i^{-1} \, D_i \qquad \textit{and} \qquad M_1 = \sum_{i=1}^{N} D_i' \, V_i^{-1} \, cov\left(Y_i\right) V_i^{-1} \, D_i'$$

$$D_i = \frac{\partial \mu}{\partial B_i'} = \begin{bmatrix} \frac{\partial E[Y]_1}{\partial \beta_1} & \cdots \frac{\partial E[Y]_1}{\partial \beta_k} & \cdots \frac{\partial E[Y]_1}{\partial \beta_m} \\ \vdots & \vdots & \vdots \\ \frac{\partial E[Y]_j}{\partial \beta_1} & \cdots \frac{\partial E[Y]_j}{\partial \beta_k} & \cdots \frac{\partial E[Y]_j}{\partial \beta_m} \\ \vdots & \vdots & \vdots \\ \frac{\partial E[Y]_p}{\partial \beta_1} & \cdots \frac{\partial E[Y]_p}{\partial \beta_k} & \cdots \frac{\partial E[Y]_p}{\partial \beta_m} \end{bmatrix}$$

The solution to the score is evaluated iteratively using the Fisher algorithm: [15]

$$\hat{\beta}^{t+1} = \hat{\beta}^{t} + \left[\sum_{i=1}^{N} D_i' V_i^{-1} D_i\right]^{-1} \times \left[\sum_{i=1}^{N} D_i' V_i^{-1} \left(y_i - \mu\right)\right]$$

D_i is a p x m matrix of the derivative of the mean at each time with respect to each parameter. A_i is a diagonal matrix of marginal variances of Y_i. As a working covariance matrix V_i should be close to the true Σ for efficiency of the estimation procedure (the score of beta is consistent irrespective of V_i)[2]. R_i is a parameterized correlation matrix with 1 on the main diagonal and the parameter vector α, representing the correlation among repeated measures, everywhere else. The most commonly used form of correlation structure is called "exchangeable", where the correlation among any pair of observations is assumed to be constant. The value of alpha is treated as a nuisance in estimating the coefficients. Several common correlation structures can be defined. "Unstructured" correlation occurs when any pair of repeated observations is assumed to have a unique value. "Auto-regressive" correlation exists when the longitudinal measurements show a steadily increasing or decreasing trend. The correlation between successive observations is modeled on the previous response. Notice that if R_i is the identity matrix then the score equation reduces to the quasi-score of a marginal GLM model which assumes independence of the repeated measures[11].

$$S\left(\hat{\beta}\right) = \sum_{i=1}^{N} D_i' \left(A_i^{\frac{1}{2}} I_{n_i} A_i^{\frac{1}{2}} \right)^{-1} \left(y_i - \mu\right) = 0$$

For the linear model, a test of $\beta_k = 0$ is a test of the effect of the covariate on the outcome of the response. The Z-statistic is used to test the significance of the coefficients.

$$Z = \frac{\hat{\beta}_k}{se\left(\hat{B}_k\right)} \sim N\left(0,1\right)$$ [1 , p.578]

GEE, as applied to normally distributed random variables, will be applied to the continuous hepatic responses. A test of the effect of the NAC treatment on the hepatic response is a test of the significance of the coefficient associated with the NAC indicator variable. The effect of NAC on each individual hepatic response will give evidence of the effect of NAC on the liver as a whole. The hepatic responses are associated with cell damage. If the effect of NAC decreases the response then NAC will be shown to be protective against cell damage induced by oxidative stress.

Suppose the longitudinal response is discrete rather than continuous. In the discrete case the parameter estimates of the covariate effects are conveniently handled in a similar manner. The only change comes from defining an alternate link function to cope with the discrete response. Suppose there is a response defining inclusion into a group of interest, the response being "yes" or "no". When dealing with binary or categorical response a logit link is frequently used. Let Y_{ij} ~Bernoulli(p) then:

$$E\left[Y_{ij}\right] = p$$

$$logit\left(\hat{p}\right) = \ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = \sum_{k=1}^{m} \beta_k \, X_k$$

The logit function is the log of the odds of being in the specified group of interest: It is assumed to be a linear function of the covariates. The expected risk associated with the outcome of interest based on this model is:

$$\hat{p} = \frac{e^{\sum_{k=1}^{m} \beta_k \, X_{ik}}}{1 + e^{\sum_{k=1}^{m} \beta_k \, X_{ik}}}$$

The estimate of coefficients through the solution to the score equation remains the same.

Likewise, suppose the discrete response is a count and assumed to have a poisson distribution with parameter lambda. In this count response case the log link is frequently used to model the covariate effects:

$$Y_{ij} \sim Poisson\,(\lambda)$$
$$E\left[Y_{ij}\right] = \lambda$$
$$\ln\left(\hat{\lambda}\right) = \sum_{k=1}^{m} \beta_k X_{ik}$$

The expected count depends on the covariate vector by the inverse transformation given below:

$$\hat{\lambda} = e^{\sum_{k=1}^{m} \beta_k X_k}$$

As mentioned previously the parameter estimates are obtained using the Fisher algorithm to solve the score equation.

Now, suppose the longitudinal response is distributed as a mixture of a poisson distribution and a bernoulli distribution. In this situation the response has some probability of coming from the bernoulli distribution and the conjugate probability of coming from the poisson distribution. Zero-Inflated Poisson (ZIP) distributions are used for this split distributed type of response. ZIP models are common in situations where the investigator is interested in deviations from a perfect state of zero response [6]. There are currently few methods to model longitudinal discrete responses that display a ZIP distribution described by Lambert [1]. A random variable Y is said to follow a ZIP distribution if Y has the following probability mass function:

$$P\left(Y = y\right) = \begin{cases} p + \left(1-p\right)e^{-\lambda}, y = 0 \\ \left(1-p\right)\dfrac{e^{-\lambda}\lambda^{y}}{y!}, y = 1, 2, 3, \ldots, 0 \le p \le 1 \end{cases}$$

The density above consists of two parts. Lam et. al. describes the first part, where y = 0, as the structural portion of the density. This is where the outcome is in a "perfect" state of zero response. (1-p) is the mixing probability or "weight" of the second portion called the "sampling" portion of the density[8]. Modeling mixed distributions pose problems because there are parameters of two different distributions associated with one random variable. If Y is distributed ZIP then:

$$Y = \begin{cases} 0, \sim Bernoulli\,(p) \\ y, \sim Poisson\,(\lambda)\ with\ mixing\ probability\ \left(1-p\right),\ y \ge 1 \end{cases}$$

Theoretically the GEE estimation procedure is applicable for models of mixed distribution, specifically models of ZIP responses. The GEE approach will be used to model the ARF scores. Intuitively the renal response of ARF is expected to be a mixture of a binary response as well as

a count response. Patients who are robust will not exhibit any risk of renal failure over the entire one year study time. They will be apart of the "structural" zero state of the response. Robust patients will remain outside of the group that displays poisson counts. Patients who develop risk following the OLT surgery are expected to fluctuate among the ARF levels of risk which include responses of zero risk at some times but could also end in ESRD. These frail patients will be apart of the sampling portion of the ZIP distribution. The treatment effect of NAC on renal dysfunction will be elucidated by assuming the ARF scores are distributed ZIP and modeling the mean ARF score as a function of patient characteristics.

2.0 APPLICATION OF GEE TO HEPATIC RESPONSE

The continuous repeated measurements of ALT, AST, Bilirubin, and INR are supposed to be correlated for each individual patient over time. However, the main interest lies in modeling the mean of each individual hepatic response and the manner in which the covariate patient characteristics, including treatment group, affect the mean response. Modeling the effect of NAC on the four hepatic responses will give an indication of the protective effect of NAC following OLT surgery. The only condition of the GEE approach is to specify the correct univariate marginal distribution of the response. Specifying the marginal distribution is accomplished by identifying the correct distribution of each response at each time-point. A "working assumption" of the correlation structure of the response is ideal since the correlation over time is treated as a nuisance. For hepatic responses the assumed correlation among any two measurements is constant. This assumption of constant correlation is used for two reasons. First, since the time intervals are not identical it would be difficult to assume that the correlation follows a regular increasing or decreasing pattern over the measured time-points. Second, the purpose of GEE is to treat this correlation as a nuisance to obtain estimates of the parameters. Using a single parameter correlation structure leads to efficiency and simplicity of the parameter estimations.

2.1 PRELIMINARY HEPATIC RESPONSE INVESTIGATION

As table 1 indicates, the patient characteristics do not differ by treatment group. If this was the case then the patient characteristic may potentially confound the actual results. For example, if a higher proportion of older patients were found in the treatment group then biased main effects would result from the model. The older patients may have worse responses due to their age which would hide the treatment effect of NAC. The Wilcoxon rank-sum test was used to compare the covariates for the two treatment groups. The Wilcoxon test is an ideal test of group differences for data of unknown distributions. Note that the p-values associated with each of the tests are non-significant at up to $\alpha = 0.12$. The treatment groups are found to be similar with respect to the patient characteristics. Males constituted 70% of the total patient population while females made up 30% of the total. Of the treatment group 70% were male. Similarly of the placebo group 69% were male. The same proportion of each gender existed in each treatment group.

Table 1: Covariate Group Comparison

Patient Characteristic	**NAC**			**Placebo**			
	P50	**P25**	**P75**	**P50**	**P25**	**P75**	***p-value**
Age	61	52	67	60	53	69	0.8453
RBC	8	5	11	6	4	9	0.1366
Don. Age	58	40	77	51	40	70	0.4646
Don. Sodium	146	141	155	146	142	153	0.8347
BMI	29.21	25.98	33.62	27.21	24.29	32.29	0.1255

* p-value is a result of the Wilcoxon Ranksum test

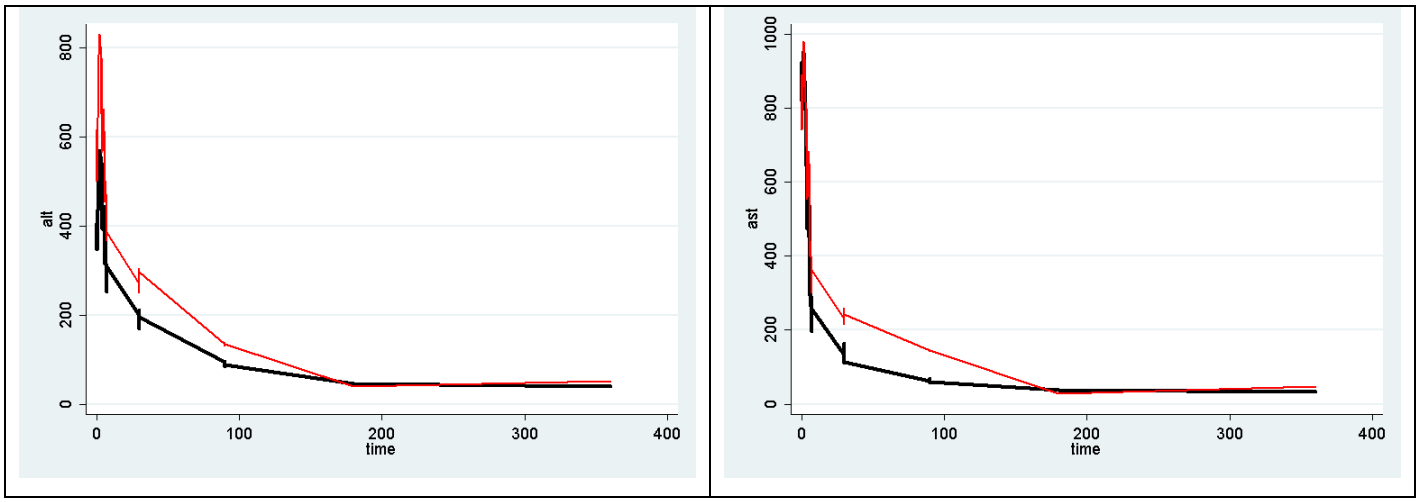

Figure 1: Lowess smooth over days of ALT and AST

ALT is the left panel AST is the right panel NAC group in red (lighter)

A locally weighted smooth-plot (lowess) was applied to all four hepatic responses. The lowess method uses a first degree polynomial least squares linear fitting method to smooth the response data over time. The bandwidth of the plots is 0.8. The bandwidth dictates the amount of nearby data used in fitting the line. These smooth plots give an indication of the average trend of the hepatic responses by treatment group over the study time. Figure 1 above depicts smooth plots of ALT in the left panel and AST in the right panel over days. The NAC treatment group is in red (lighter). It seems as though the NAC group initially exhibits increased levels of both ALT and AST. However, as time passes the ALT and AST of both treatment groups equalize starting at 180 days from the surgery. The graphs of Figure 2 below depict the lowess smooth plots of bilirubin on the left and INR on the right. The bilirubin level for the NAC group is consistently above the placebo group throughout the entire study time. It seems as though NAC increases bilirubin levels following OLT surgery. The INR response is interesting. The NAC group starts with longer blood clotting times, but by the end of the study the situation is reversed. The placebo group seems to have longer clotting times compared to the treated group at the end of one year.

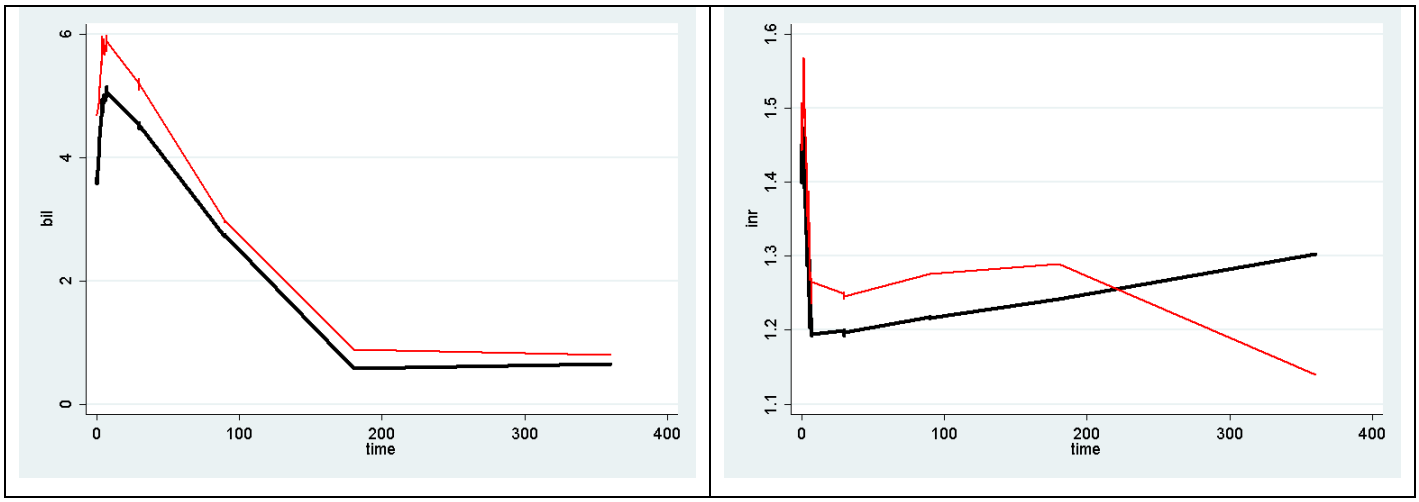

Figure 2: Lowess smooth over days bilirubin and INR

Bilirubin is the left panel INR is the right panel NAC group in red (lighter)

The distributions of all four raw hepatic responses are skewed over all time-points as shown by Figures 3-6. In order to comply with the assumption of normally distributed responses the raw measures must assume a transformation. The natural log transformation is used for the hepatic responses of ALT, AST, and bilirubin. The reciprocal square root transformation is chosen for the response of INR. Side by side histograms over time are displayed in Figures 3-6. These figures show substantially more evidence for normally distributed measures after the respective transformations. The transformed hepatic responses are systematically related to a linear function of the covariates. Estimates of the effects of patient characteristics, including treatment status can be determined through an identity link with the response.

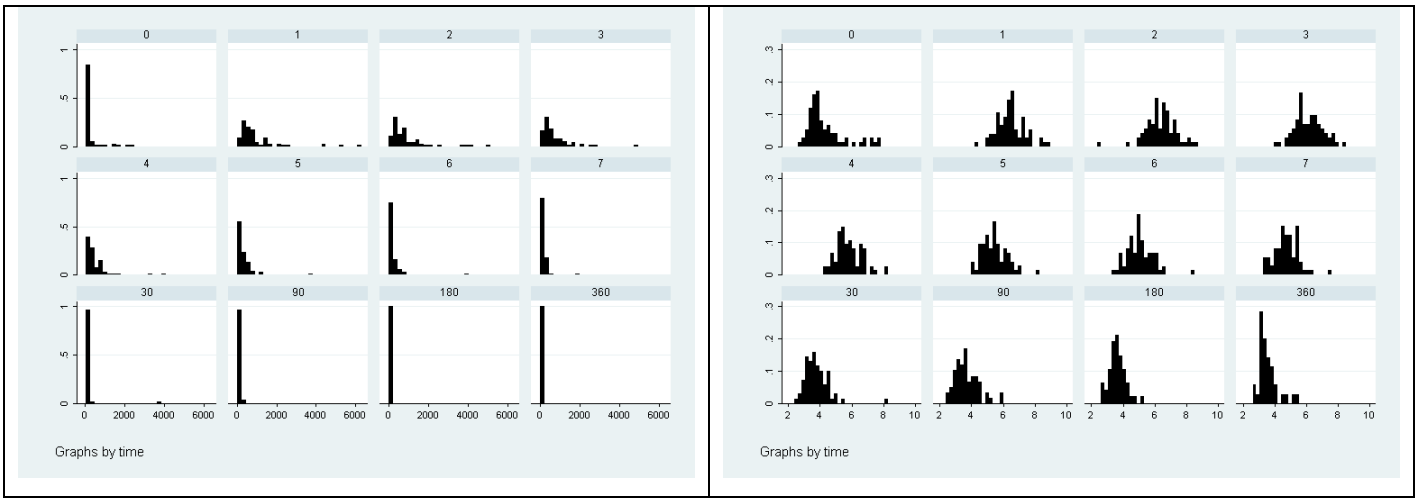

Figure 3: ALT histograms over days

The distributions are shown over days the left panel contains the raw ALT and the right panel contains the log transformed ALT

Figure 3 displays the transition in the distributions over time of ALT before and after the log transformation. The hepatic response of ALT does not seem normally distributed before transformation. The left panel shows histograms over time of the raw ALT response, while the panel on the right consists of histograms of the natural log transformed response over the measured times.

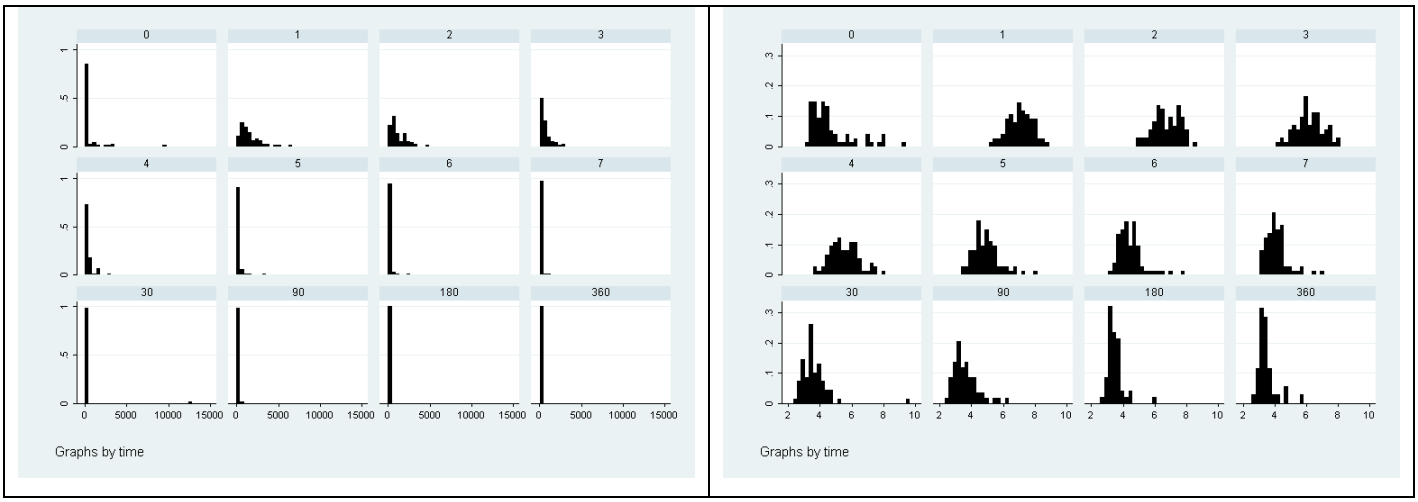

Figure 4: AST histograms over days

The distribution of AST is shown over days the left panel contains the raw AST while the right panel contains the log transformed AST

Figure 4 is a comparison of the distributions over time of AST before and after the log transformation. There is weak evidence for the normality assumption of AST before the natural log transformation. The left panel consists of histograms of the raw AST response over time. The right panel is that of the natural log of AST over time.

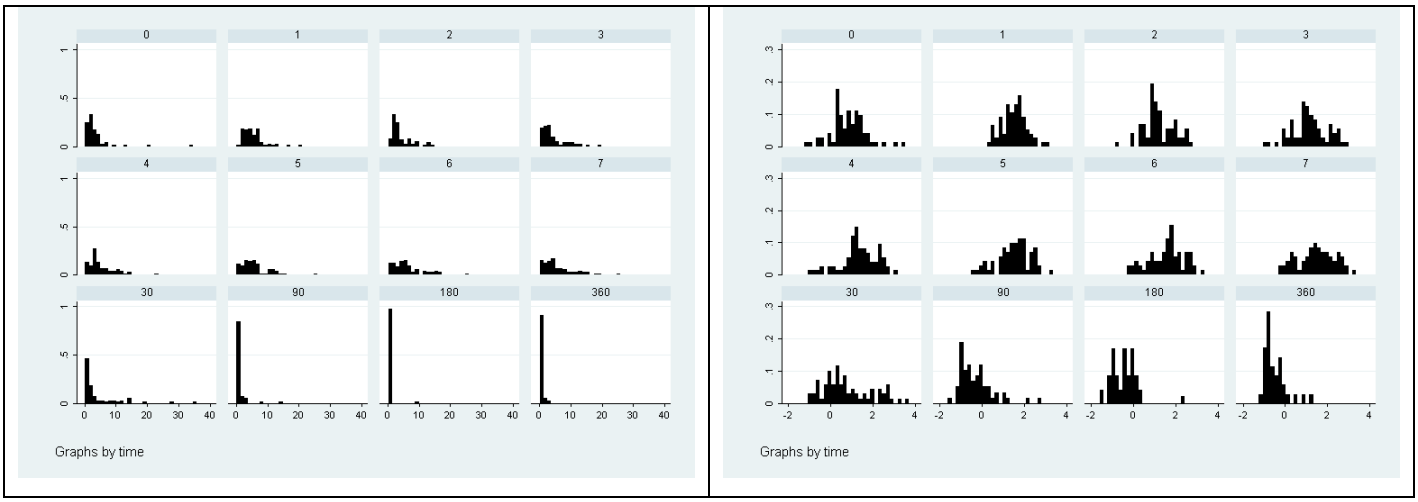

Figure 5: Bilirubin histograms over days

The distribution of bilirubin over days the left panel is that of the raw bilirubin the right panel is the log transformed bilirubin.

Figure 5 compares the bilirubin measurements over time before and after a log transformation. The distribution of Bilirubin seems skewed to the right before the transformation is applied. After the transformation, there is strong visual evidence supporting the assumption that the response is normally distributed over time.

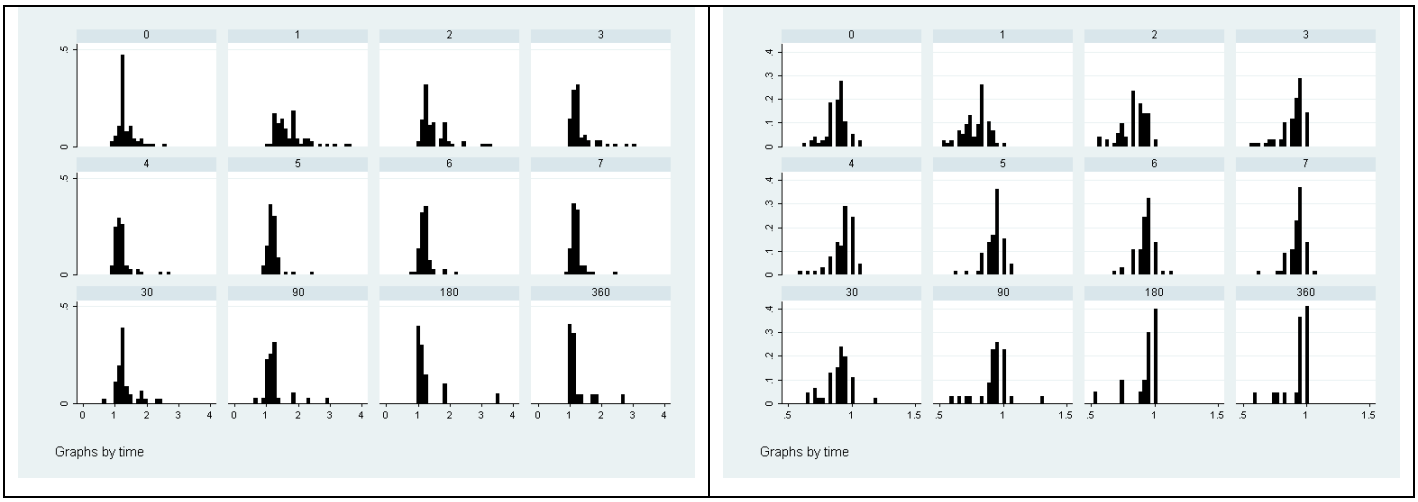

Figure 6: INR histograms over days

INR distribution over days left panel is the raw INR level the right panel is the reciprocal square root transformation

Figure 6 depicts the transition in distributions over time of INR before and after the reciprocal square root transformation. The hepatic response of INR is skewed toward lower values initially. Following the reciprocal square root transformation the INR values are more normally distributed. In summary, transformations for the responses are as follows:

$$E\left(Y_{ij}\right)=\sum_{k=1}^{m} B_k X_{ik}$$

$$where\, Y_{ij} = \begin{cases} \ln(ALT) \\ \ln(AST) \\ \ln(Bilirubin) \\ INR^{-\frac{1}{2}} \end{cases}$$

Once the model fitting is completed, to revert back to the original scale an inverse transformation must be made. When an estimate of the response is calculated for a patient, the inverse function is chosen to transform the $E(Y_{ij})$ back to the original scale of the response. If Z is the original scale then

$$Z_{ij} = \exp\left(Y_{ij}\right) \; if\; Y_{ij} = \ln(ALT); \ln(AST); \ln(Bilirubin)$$

$$Z_{ij} = Y_{ij}^{-2} \; if\; Y_{ij} = INR^{-\frac{1}{2}}$$

In modeling the mean response for each of the four hepatic models, the continuous covariates are centered on the mean. Centering the continuous covariates is done for two reasons. First, the main goal of the analysis is to assess the effect of the treatment NAC on the response. Second, centering the covariates makes sense for direct comparisons to an average patient. In the population averaged model the baseline patient is a female who is not treated with average age, average amount of blood transfused, average donor serum sodium level, average donor age, and average BMI of the patient population.

Table 2: Means of Patient Characteristics

Patient Characteristic	Mean
Age	59.98
Blood Trans.	7.52
Don. Sodium	146.74
Don. Age	55.25
BMI	29.13

Table 2 shows the means of the patient characteristics. The average OLT patient is 60 with a BMI of 29.13. The average transfused blood received during surgery is 7.52. The average donor serum sodium and age are 146.74 and 55.25 respectively.

2.2 NORMAL GEE MODELS FOR HEPATIC RESPONSES.

Assuming the log transformed ALT is normally distributed and using an identity link for the GEE estimation procedure produces results listed in Table 3 . Estimates were computed using the *xtgee* command in STATA version 9.2. The unadjusted estimates of Table 3 are a result of modeling a single covariate at a time to evaluate the effect of the covariate on the hepatic response. The adjusted model shown in Table 4 contains only the covariates that were individually significant in the unadjusted model. Estimates of the regression coefficients of the adjusted multi-variable model are obtained simultaneously. For the GEE approach there are limited variable selection methods[1]. The only two variables retained from the univariate case are patient age, treatments status, and time. Although the effect of time on ALT is small it needs to be adjusted for in the final model. Treatment status was not significant according to the listed p-value, however it is the main focus of the analysis. NAC increase the ALT as shown by the positive coefficient. However, there is weak evidence for this association as indicated by the respective p-value of 0.207. ALT is a measure of hepatic cell death, and an increase is unusual for the assumed protective effect of NAC.

Table 3: Log ALT GEE model unadjusted estimates

Unadjusted Estimates

Variable	Coeff.	95% C.I.	P-Value
NAC	0.157	(-0.1086 , 0.4225)	0.247
Age	-0.020	(-0.0336 , -0.0071)	<0.01
Male	-0.088	(-0.4119 , 0.2354)	0.593
Blood Trans.	-0.013	(-0.0305 , 0.0035)	0.120
Don. Sodium	0.008	(-0.0053 , 0.0222)	0.230
Don. Age	0.002	(-0.0056 , 0.0097)	0.594
BMI	0.009	(-0.0128 , 0.0310)	0.416
Days	-0.007	(-0.0080 , -0.0063)	<0.01

Table 4: Log ALT GEE model adjusted estimates

Adjusted Estimates

variable	Coeff.	95% C.I.	P-value
NAC	0.156	(-0.0865 , 0.3988)	0.207
Age	-0.015	(-0.0276 , -0.0017)	0.026
Days	-0.007	(-0.0079 , -0.0062)	<0.01
Const.	5.285	(5.1258 , 5.4447)	<0.01

The model for ALT is as follows:

$$Ln\,(ALT) = 5.285 + treatmentgroup \times 0.156 - \left(age - 60\right) \times 0.018 - days \times 0.007$$

Therefore the average 60 year old OLT patient who was treated with NAC has estimated ALT at one week following surgery of exp(5.285 + 0.156 – 0.049) = 219.64. On the other hand the estimated ALT for an average 60 year old patient on placebo is exp(5.285 – 0.049) = 187.91. The estimated difference of ALT between the treatment and placebo groups is not substantial. The age coefficient is negative. Being older than average decreases the level of estimated ALT.

The information listed in Table 5 is the result of modeling the natural log of AST as a linear function of covariates using the GEE procedure. The unadjusted estimates are the result of univariate models as was the case with the ALT response. Patient age is the only covariate that may aide in determining the outcome of AST level. Table 6 lists the results of the full AST model adjusted for the significant age variable, as well as treatment group and time. According to Table 6 there is no evidence that NAC has a significant effect on the outcome of AST as shown by the p-value of 0.740.

Table 5: Log AST model unadjusted estimates

Unadjusted Estimates

Variable	Coeff.	95% C.I.	P-value
NAC	0.0472	(-0.200 , 0.294)	0.708
Age	-0.011	(-0.024 , -0.0015)	0.083
Male	-0.125	(-0.491 , 0.169)	0.405
Blood Trans.	-0.0041	(-0.0186 , 0.0104)	0.583
Don. Sodium	0.0092	(-0.004 , 0.0231)	0.195
Don. Age	0.0020	(-0.004 , 0.0086)	0.55
BMI	0.0142	(-0.006 , 0.034)	0.176
Days	-0.007	(-0.0084 , -0.0068)	<0.01

Table 6: Log AST model adjusted estimates

Adjusted Estimates

Variable	Coeff.	95% C.I.	P-value
NAC	0.041	(-0.2008 , 0.2827)	0.740
AGE	-0.006	(-0.0192 , 0.0076)	0.394
Days	-0.007	(-0.0083 , -0.0067	<0.01
Const.	5.245	(5.0888 , 5.4013)	<0.01

The effect of age on AST also shows a reduced association in the fully adjusted model compared to the univariate case. The p-value changed from 0.083 in the univariate model to 0.394 in the fully adjusted model. The change of significance is due to the adjustment for time. NAC is proposed to have an increasing effect on AST although the effect is close to zero. An increase in age causes the AST level to decrease. Like the treatment effect, the amount of decrease in AST due to age is close to zero. The full form of the AST model is as follows:

$$Ln\,(AST) = 5.24 + treatmentgroup \times 0.041 - \left(age - 60\right) \times 0.006 \; - \; days \times 0.007$$

The average aged patient on treatment at one week post-op has and estimated AST level equal to exp(5.24 + 0.041 - 0.049) = 187.16. The placebo patient at the same time has an almost identical estimated AST level of exp(5.24 – 0.049) = 179.65. Since ALT and AST are closely related it is surprising that NAC would show a stronger association with ALT than with AST. There may be some hidden confounding variable affecting the outcome. Alternatively the mechanism of action of the treatment on these proteins may not be fully understood.

The unadjusted estimates of covariate effects for bilirubin level are given in Table 7. The covariates of blood transfusion, donor serum sodium level, and donor age are significant as shown by the respective p-values of 0.06, 0.03, and 0.09. The p-value of NAC is not significant at 0.388. In the adjusted model of Table 8 there is no dramatic change in the strength of the association of the covariates.

Table 7: Bilirubin model unadjusted estimates

Unadjusted Estimates

Variable	Coeff.	95%	P-value
NAC	0.127	(-0.162 , 0.418)	0.388
Age	-0.002	(-0.017 , 0.012)	0.775
Male	-0.191	(-0.533 , 0.150)	0.272
Blood Trans.	0.0195	(-0.0008 , 0.0398)	0.06
Don. Sodium	0.0168	(0.0011 , 0.0326)	0.036
Don. Age	0.007	(-0.0011 , 0.015)	0.092
BMI	0.0182	(-0.007 , 0.0435)	0.158
Days	-0.006	(-0.0072 , -0.0054)	<0.01

Table 8: Bilirubin model adjusted estimates

Adjusted Estimates

Variable	Coeff.	95% C.I.	P-value
NAC	0.074	(-0.1759 , 0.3243)	0.561
Blood Trans.	0.019	(0.0048 , 0.0330)	<0.01
Don. Sodium	0.015	(0.0028 , 0.0267)	0.016
Don. Age	0.008	(-0.0012 , 0.0156)	0.022
Days	-0.006	(-0.0072 , -0.0054)	<0.01
Const.	1.23	(1.0312 , 1.4385)	<0.01

As one would expect donor age, donor serum sodium levels and amount of blood transfused have increasing effects on the level of Bilirubin. In contrast to the hypothesized protective effect of the treatment on liver function, NAC increases the level of bilirubin, although not significantly. As in the model of AST, the evidence for significance of NAC in affecting the Bilirubin level is very weak with a p-value of 0.527. The model for bilirubin is as follows:

$$Ln\,(Bilirubin) = 1.23 + treatmentgroup \times 0.074 + \left(Blood\,transfused - 7.52\right) \times 0.019 +$$
$$(Don \cdot Sodium - 146.74) \times 0.015 + \left(Don \cdot Age - 55.25\right) \times 0.008 - days \times 0.006$$

An estimate of bilirubin level for a patient from the treatment group with average blood transfused, average donor serum sodium, and average donor age at one week following the surgery is exp(1.23 + 0.074 – 0.042) = 3.52. For the same average patient from the placebo group the estimate is exp(1.23 – 0.042) = 3.28, which is nearly identical to the treated patient. NAC does not show an association with the level of bilirubin post-transplant.

Tables 9 and 10 give the unadjusted and adjusted model estimates for the response of INR. INR is a measure of blood coagulation time. The transformation used to assume a normally distributed response is the reciprocal square root. In this case a positive coefficient will decrease the INR, and a negative coefficient will cause the estimate of INR to increase. This relationship is due to the inverse transformation used to obtain the raw INR level. According to Table 9, the variables NAC patient age, and time are significant. In the fully adjusted model of Table 10 the strength of the association of the treatment to the INR outcome is strengthened.

The p-value changes from 0.077 in the univariate case to 0.043 in the adjusted case. The strength of the age association decreases slightly.

Table 9: INR model unadjusted estimates

Unadjusted Estimates

Variable	Coeff.	95% C.I.	P-value
NAC	-0.022	(-0.0467 , 0.0024)	0.077
AGE	0.002	(0.0004 , 0.0032)	0.010
Male	-0.020	(-0.0474 , 0.0060)	0.130
Blood Trans.	-0.001	(-0.0026 , 0.0007)	0.268
Don. Sodium	-0.001	(-0.0025 , 0.0003)	0.128
Don. Age	-0.0003	(-0.0009 , 0.0003)	0.330
BMI	-0.001	(0.8732 , 0.8981)	0.283
Days	0.0001	(0.0 , 0.0003)	0.03

Table 10: INR model adjusted estimates

Adjusted Estimates

Variable	Coeff.	95% C.I.	P-value
NAC	-0.023	(-0.0465 , -0.0007)	0.043
Age	0.002	(0.0004 , 0.0032)	0.013
Days	0.00015	(0.0 , 0.0003)	0.029
Const.	0.893	(0.8757 , 0.9110)	<0.01

The model of INR is as follows:

$$INR^{-\frac{1}{2}} = 0.893 - treatmentgroup \times -0.023 + \left(age - 60\right) \times 0.002 + days \times 0.00015$$

so for a treated patient of average age 60 at one week following OLT surgery the estimated INR is $(0.893 - 0.023 + 0.001)^{(-2)} = 1.318$. For a similar untreated patient the INR at one week is $(0.893 + 0.001)^{(-2)} = 1.251$. The treatment increases the time for blood coagulation. Increased blood coagulation time is an indication of hepatic cell damage. Increases in age result in decreases in INR and therefore decreases in clotting time. This decrease in coagulation time with respect to age is suspect. The contradictory evidence gives further support to a possible unknown confounding variable.

It is interesting that the treatment NAC is shown to increase these indicators of hepatic cell damage. There is weak evidence supporting the contention that the treatment is associated with levels of ALT, AST and bilirubin in the post-transplant period of one-year. As supported by the respective p-values there is moderate evidence that the treatment is associated with increased levels of the prothrombin time. The increase in prothrombin time by NAC may be a result of increased cell damage. There could possibly be an unknown confounding variable that sits between the treatment effect and the mechanism for blood coagulation. There may be an additional unknown hierarchical grouping variable associated with the hepatic function following OLT. For example donor organ characteristics may play a larger role then assumed in determining hepatic functioning level. Further analysis may need to include more information on donor characteristics.

3.0 APPLICATION OF ZIP GEE TO RENAL RESPONSE

The distribution of ARF scores seems to be distributed as Zero Inflated Poisson at each time point because of the large proportion of zero scores at each time. At the first time-point, which occurs prior to surgery, all patients have an ARF score of zero. Initially, each patient does not exhibit any risk of kidney failure. On the first day following OLT surgery only seven patients have scores designating some renal failure risk. As days pass more patients switch from the structural portion of the model to the sampling portion. The structural zeros can be thought to be distributed as Bernoulli with probability (p) of the ARF score having the value zero. Patients who are relatively healthy and resilient will have zero scores over all times and will not enter the sampling portion of the distribution. Frail patients, on the other hand, transition from the zero state to the non-zero state. They will fluctuate among all ARF levels during the study period.

3.1 ARF DATA

ARF scores measure the condition of the patient's renal functioning status. As characterized by the RIFLE criteria.[7] The ARF score gives an indication of the damage done to the kidneys following liver transplantation. Table 11 is a list of the proportion of zero ARF scores at each measured time-point. The proportions of patients who have no risk of renal failure seem to decrease over time. More patients transition from the zero-state and become part of the non-zero state as time passes.

Table 11: Proportion of zero ARF at each time

p	estimate
P1	.999
P2	.905
P3	.819
P4	.704
P5	.685
P6	.671
P7	.766
P8	.625
P9	.555

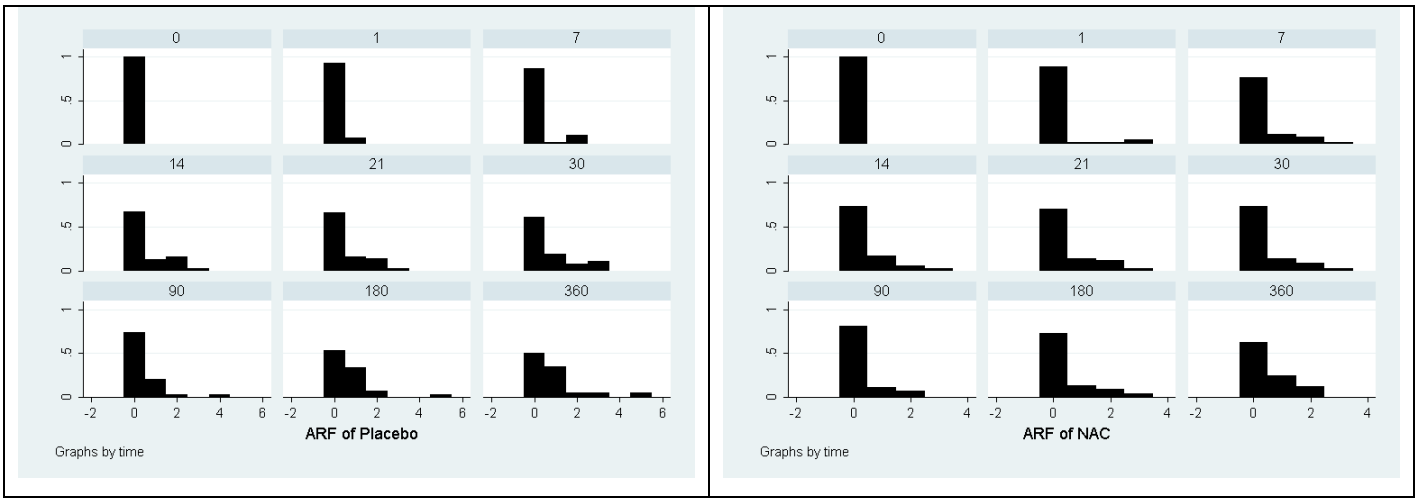

Figure 7: Distribution of ARF by treatment group over days

Placebo group is in the left panel while treatment group is in the right panel

The distribution of ARF scores over time in days is shown in Figure 7. The left panel consists of the placebo group. The panel on the right shows the treatment group. There is a large proportion of zero ARF scores at each time-point. While the non-zero values that denote some risk of renal failure are sparse and fluctuate over all times. Also note the small proportion of ones in the treatment group at the ending three time-points compared to the proportion of ones in the placebo group at the same times.

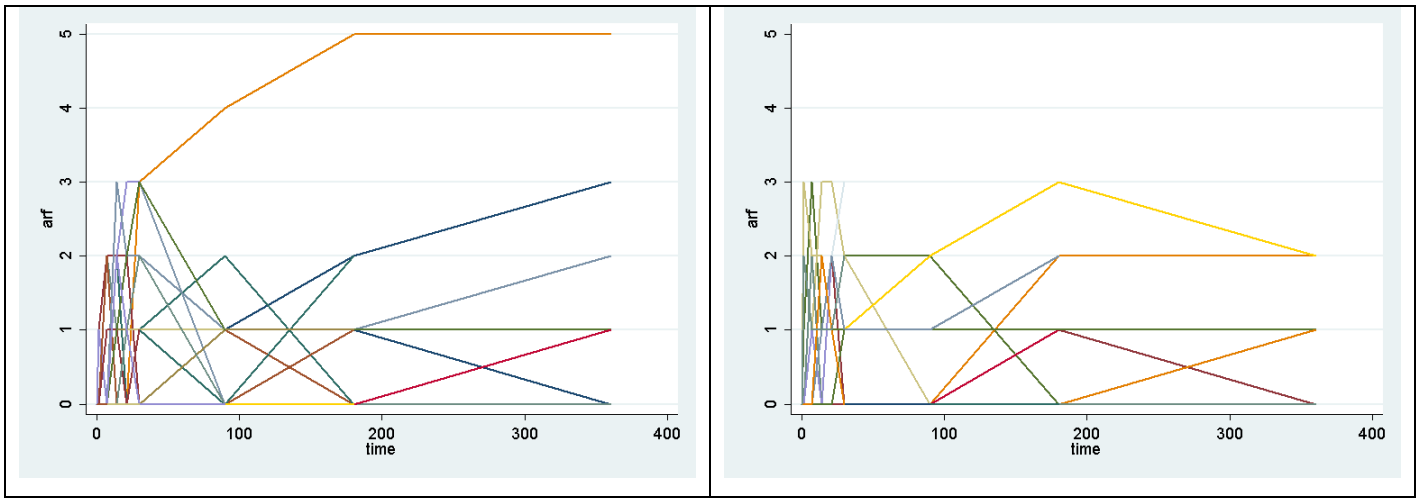

Figure 8: Spaghetti plots ARF over days

The panel on the left depicts the placebo group the panel on the right depicts the treatment group

A closer look at the conditions of each individual patient over time is given in the graphs of Figure 8 and 9. Figure 8 depicts spaghetti plots of each individual patient. The graphs are connected line plots of the ARF scores for each patient over the one-year study time. The scores of the treatment group have a maximum of 3. On the other hand, the scores for the placebo group not only surpass 3 but seem to increase over the study time compared to the treatment group. Figure 9 depicts the fitted values of ARF score over time. Again, the trend seems to be increasing in the placebo group and decreasing in the control group. NAC seems to be decreasing the ARF score at each time-point. There is substantial graphical evidence that NAC has a protective effect on the kidneys following OLT surgery.

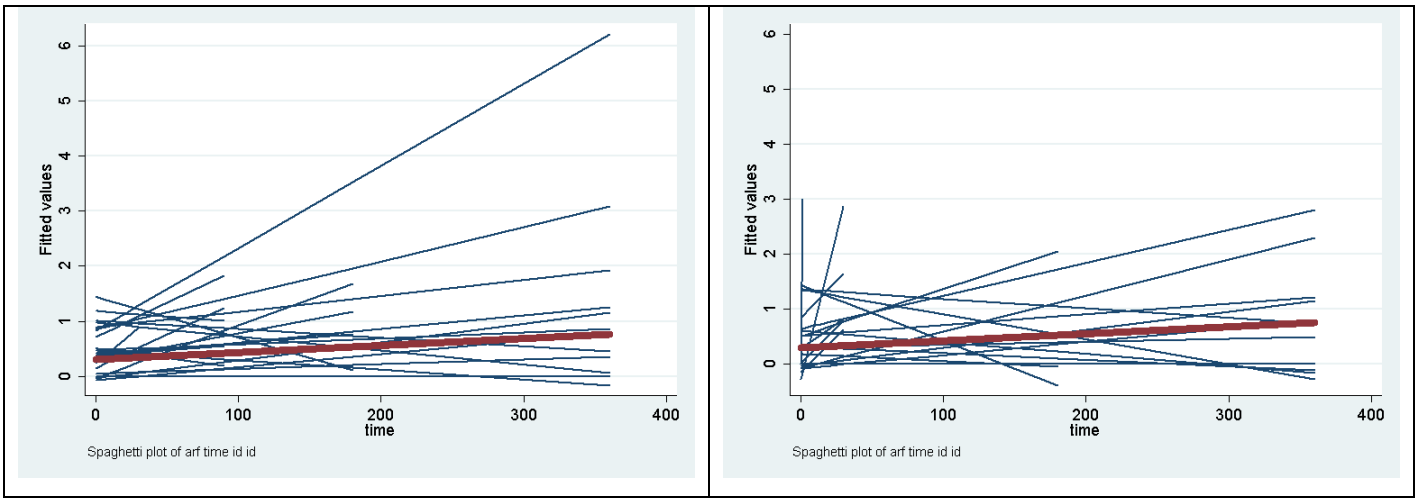

Figure 9: Fitted Values ARF over days

The panel on the left is the placebo group the panel on the right is the treatment group

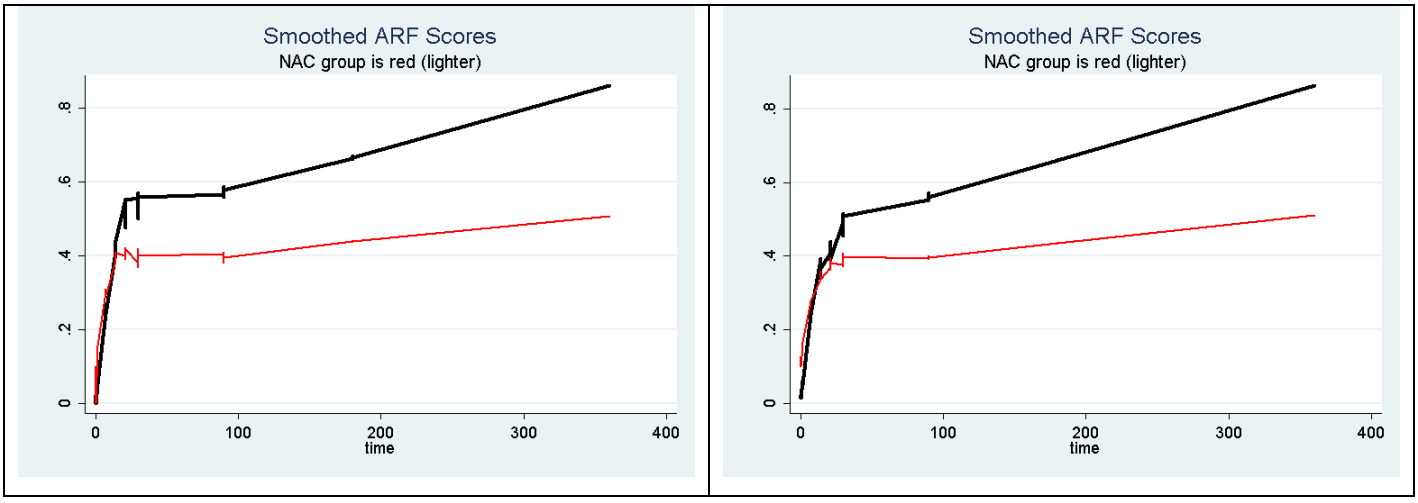

Figure 10: Smoothed ARF score over days by treatment group

The left panel uses bandwidth of 0.6 right panel uses bandwidth of 0.8

A locally weighted smooth-plot (lowess) was applied to the ARF scores. The lowess method uses a first degree polynomial least squares linear fitting method to smooth the discrete response data over time. The treatment group shows a marked difference in ARF scores. The difference between the treatment and control groups is supported by this smooth plot. The

treatment group exhibits lower ARF scores compared to the placebo group. Notice that ARF is highly skewed toward zero at the first time as one would expect from a ZIP distribution. The left graph of Figure 10 uses a bandwidth of 0.6 while the graph on the right uses a bandwidth of 0.8. The bandwidth dictates the amount of nearby data used in fitting the line.

3.2 ZIP MODEL FOR ARF SCORE

Initial parameter estimates are obtained by fitting the GEE model using the natural log link of the Poisson distribution for ARF. Molenbergh[15] advises to start with univariate estimates in the Fisher algorithm. The Poisson model does not take into account the high proportion of zero scores at each time-point. Initial estimates of the p_j at each time-point are taken to be the proportion of zero ARF at each time as listed in Table 11. Table 12 gives the initial estimates obtained through Poisson GEE.

Table 12: Initial unadjusted model estimates for ARF GEE

Variable	Coeff.	P-value
NAC	-0.146	0.612
Age	-0.026	0.155
Male	0.070	0.830
Blood Trans.	-0.040	0.211
Don. Sodium	0.027	0.082
Don. Age	-0.006	0.336
BMI	0.010	0.673

As mentioned in Section 1.3 The ARF response is assumed to be distributed as:

$$Y_{ij} = \begin{cases} 0\ , \sim Bernoulli\left(p_j\right) \\ y, \sim Poisson\left(\lambda\right) \ with\ probability\ \left(1-p_j\right)\ ,\ y \geq 1 \end{cases}$$

For a ZIP distributed random variable as given by [9]:

$$E\left[Y\right] = \left(1-p\right)\lambda$$

$$Var\left[Y\right] = \lambda\left(1-p\right)\times\left[1+\lambda p\right]$$

Modeling the ZIP distributed longitudinal response as a function of covariates is accomplished through the GEE method previously described. The mean is modeled through component parameters as follows:

$$E\left[Y_{ij}\right] = \left(1 - p_j\right)\ \lambda$$

Where p is modeled as a logit regression on time-dependent intercepts:

$$\hat{p}_j = \frac{e^{\gamma_j}}{1 + e^{\gamma_j}}$$

In order to test the association of covariates with emphasis on the treatment, the parameter lambda is modeled as an exponential form of the covariate vector X. To introduce the covariate patient characteristics: The parameter lambda is modeled as a time-independent function of covariates given by:

$$\hat{\lambda} = e^{\sum_{k=0}^{m} X_{ik}\,\beta_k}$$

The expectation and variance of the model is then given by:

$$E\left[Y_{ij}\right] = \left(1 - p_j\right)\ e^{\sum_{k=1}^{m} X_{ik}\,\beta_k}$$

$$Var\left[Y_{ij}\right] = \ e^{\sum_{k=1}^{m} X_{ik}\,\beta_k}\left(1 - p_j\right) \times \left[1 + \ e^{\sum_{k=1}^{m} X_{ik}\,\beta_k}\ p_j\right]$$

The covariates are only involved in the poisson state. Since the ZIP model has a defined marginal mean the GEE approach can be applied. The Fisher algorithm is used to solve the score as in the general case. MATLAB was used to evaluate the modified Fisher algorithm proposed by Liang and Zeger [10]. For the MATLAB code refer to Appendix A. The exchangeable correlation structure is assumed. As mentioned previously the exchangeable correlation signifies the assumption of constant correlation among all time-points. One would initially assume the correlation among time intervals is constant because each time interval is of a different size. The main goal is to understand the effect of NAC on renal functioning as measured by ARF score at each time. The correlation over time is a nuisance as is ignored to obtain efficient parameter estimates of the treatment effect. The exchangeable alpha is estimated by the following:

$$\hat{\alpha} = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{n_i\left(n_i - 1\right)} \sum_{j \neq k} e_{ij}\ e_{ik}$$

$$e_{ij} = \frac{y_{ij} - \mu_{ij}}{\sqrt{Var\left(u_{ij}\right)}}$$

3.3 RENAL ANALYSIS RESULTS

The ZIP GEE model explained above produces the following parameter estimates shown in Table 13. The p-value of the treatment effect of 0.1047 indicates a moderate association with the outcome measure of ARF score as an indicator of renal function status. Furthermore the effect of NAC as given by the coefficient of -0.303 is to reduce the ARF score. Univariate models of the other patient characteristics do not produce significant changes in ARF score. Also, the estimates of the effects are substantially smaller compared to NAC.

Table 13: ZIP GEE model for ARF score univariate estimates

Unadjusted Estimates

Variable	Coeff.	95% C.I.	P-Value
NAC	-0.303	(-0.7761 , 0.1701)	0.1047
Age	-0.0017	(-0.0272 , 0.02378)	0.446
Male	0.0117	(-0.4844 , 0.5079	0.518
Blood Trans.	-0.020	(-0.0692 , 0.0284)	0.207
Don. Sodium	0.011	(-0.0147 , 0.0372)	0.801
Don. Age	-0.003	(-0.0150 , 0.0087)	0.302
BMI	0.01	(-0.0266 , 0.0466)	0.701

Estimates of the time-dependent parameters p_j using NAC as the single covariate are given in Table 14. The proportion of zero ARF prior to surgery is omitted because it is one.

Table 14: Estimates of p_j of ZIP GEE with NAC covariate

P_j	Estimate
P1	0.8064
P2	0.6400
P3	0.4699
P4	0.4312
P5	0.3704
P6	0.5941
P7	0.3215
P8	0.2056

The estimates of the proportion of zero ARF scores computed from the ZIP GEE seem to decrease over time. One can construct a Wald statistic to test the hypothesis that the p_j's are all equal. This test will indicate whether the proportion of zero ARF scores change over time. The Wald test is defined as follows:

$$V_p = cov\left[\hat{p}_j\right]$$

$$V_p* = C \times V_p \times C'$$

where:

$$C = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

then:

$$\chi^2\left(p-1\right) = \hat{p}' \times \left(V_p*\right)^{-1} \times \hat{p}$$

The chi-square statistics is 166.46 giving a p-value of much less than 0.01. The proportions of zero ARF responses are not constant over time. There are patients who transition from the zero state and enter the non-zero state of the ZIP distributed response. The model for ARF is as follows:

$$\hat{ARF} = \left(1 - p_{time}\right) e^{treatmentgroup \times (-0.303)}$$

As an example, the expected ARF for a patient on NAC treatment at the 14th day is (1-0.4699)*exp(-0.303) = 0.3914 while the expected ARF for a patient in the placebo group at the 14th day is (1-0.4699) = 0.530. Notice that a constant is not included in the model. A constant in the above model will introduce a scale of the term $(1-p_j)$. This scale would cause the estimate of the constant to be non-identifiable from the parameter (p_j) which is the proportion of zero ARF scores. It would not be possible to separate the estimate of p_j from the estimate of the constant. The graphical analysis, as shown by the locally weighted scatter-plot smooth of Figure 10, reinforces the hypothesis that NAC is protective against renal damage during OLT surgery. This analysis warrants further investigation of the NAC as a prophylaxis for renal damage following OLT surgery.

4.0 DISCUSSION

The GEE approach was applied to the hepatic responses of OLT patients in order to elucidate the effect of the treatment NAC on liver dysfunction. NAC is thought to be protective against damage by oxidative stress following transplant surgery. The hepatic responses of ALT , AST , Bilirubin, and INR had to be transformed in order to achieve the assumption of a normally distributed response. The GEE estimates of treatment effect contradicts the hypothesis that NAC is protective against liver damage following OLT. All four indicators of hepatocyte cell damage increases in association with the treatment. There is a strong possibility that an unknown confounding variable exists which is affecting the status of liver function following surgery. One possibility is the effect of the donor liver status in the outcome of liver graft performance. The status of the donor liver would have a large influence on the functioning status of the transplant patient. Further analysis of the hepatic response should include more information concerning donor liver characteristics.

The ARF scores, as a measure of renal dysfunction following surgery, were shown to exhibit the unique Zero-Inflated Poisson distribution. Graphical analysis supported a treatment effect of NAC in protecting against renal dysfunction. Initial estimates using a log link of an assumed poisson distributed response yielded inaccurate effects of the treatment. The GEE procedure as proposed by Liang and Zeger [10] was used to estimate the true treatment effect of NAC. The GEE procedure easily accommodated the uniquely distributed ARF response. Further investigation into the effectiveness of NAC as a prophylaxis to prevent renal damage is necessary.

APPENDIX A

MATLAB PROGRAM FOR ZIP GEE

```
% %  the initial data must consist of
% %  Response Scores (ARF)
% %  Initial exchangeable correlation alpha
% %  covariate vector (nac age male donage donna bmi)
% %  Some Intial parameters the p-vector and the beta-vector


%%  Initial Estimates

alpha = .303;

p = [.9999;.9054;.8194;.7042;.6857;.6714;.7666;.6250;.5555];

gamma = log(p./(1-p));

%univariate beta estimations from
%  xtgee poisson  [-.146;-.026;.07;-.04;.027;-.006;.01]

beta = -.146;

parm = [gamma; beta(:)];

parmnew = parm;

R = alpha*ones(9);

for w=1:9;
    for z=1:9;
        if w == z;
            R(w,z)=1;
        end;
    end;
end;

% Multivariate design matrix
% desX = [nac (age-mean(age)) male (rbc-mean(rbc)) (donna-mean(donna))
```

```
%   (donage-mean(donage)) (bmi-mean(bmi))];

desX = [nac];

q = 1 ;

tol = 1;


%% global loop starts here
while (q < 500) && (tol > .0001 )

                beta = parmnew(10);

                gamma = parmnew(1:9);
     %% the following keeps p bounded in [0,1]
                p = exp(gamma)./(1 + exp(gamma));

%%  The initilization of estimation sum terms.

%  Initialize the iterative coefficient estimate sums

sumterm1 = 0*ones(10);
sumterm2 = 0*ones(10,1);

%  Initialize Standard Error Sums
Mnull = zeros(10);
Mone = zeros(10);
RobustMone = zeros(10);


%% inner loop starts here Sum over panels (patients)

for i = 1:74

% vary is a column vector of marginal variances for patient i.

vary = exp(sum(desX(i,:)'.*beta))*(1-p).*(1 + exp(sum(desX(i,:)'.*beta))*p) ;

% Marginal Co-Variance Matrix

V = sqrt(diag(vary))*R*sqrt(diag(vary)) ;


%% Creating the Expectation column vector.

EY = (exp(sum(desX(i,:)'.*beta))*(1-p)) ;


%% The following creates the Dmu matrix

f = -(exp(sum(desX(i,:)'.*beta)))*eye(9) ;
```

```
l = (desX(i,:)'.*exp(sum(desX(i,:)'.*beta)))';

m = (1-p)*l;

Dmu = [f m] ;

%% Now for the Sums over panels for parameter estimate

% the terms are for i and must be
% summed over all patients.

term1 = Dmu'*inv(V)*Dmu ;

term2 = Dmu'*inv(V)*(arf(i,:)'-EY) ;

sumterm1 = sumterm1 + term1;
sumterm2 = sumterm2 + term2;

%%   for the std. errors

mnull = Dmu'*inv(V)*Dmu;

mone = Dmu' * inv(V) * ((arf(i,:)' - EY) * (arf(i,:)' - EY)') * inv(V) * Dmu;

moneR = Dmu' * inv(V) * (cov(arf)) * inv(V) * Dmu;

Mnull = Mnull + mnull;

Mone = Mone + mone;

RobustMone = RobustMone + moneR;

end
%% End of Inner loop over panels


%% Updating the parameter estimates.

 %  Used to check the previous parameter
oldparm = parmnew;

parmnew = parmnew + (inv(sumterm1)*sumterm2);

 %  A tolerance vector
tolvec = abs(parmnew - oldparm);

 %  setting the tolerance equal to the max of any individual difference
tol = max(abs(parmnew - oldparm));

 %   Go to next iteration
```

```
q = q+1;

end

%% End of Outer loop for each updating iteration.


%% Outputting the Results of the Estimating Equations

format long g

parmnew

% Naive Beta Variance
NVparm = diag(inv(Mnull)*Mone*inv(Mnull));

% Robust Beta Variance
RVparm = diag(inv(Mnull)*RobustMone*inv(Mnull))

% Wald using Naive estimate variance
NWald = parmnew ./ (sqrt(NVparm));

% Wald using Robust estimate variance
Wald = parmnew ./ (sqrt(RVparm))

q

tol

gamma = parmnew(1:9);

%  estimates of proportion of zeros at each time-point
p = exp(gamma)./(1 + exp(gamma));

nacstderr = sqrt(RVparm(10))

nacest = parmnew(10)

nacZ = Wald(10)
```

BIBLIOGRAPHY

[1] Agresti, Alan. (2002). Categorical Data Analysis. Hoboken: John Wiley & Sons, Inc.

[2] Bromley PN, Cottam S.J, Hilmi I, et al. Effects of intraoperative N-acetylcysteine in orthotropic liver transplantation. BJA 1995; **75**: 352-354

[3] Diggle, Peter J. ; Liang, Kung-Yee ; Zeger, Scott L. (1994). Analysis of Longitudinal Data. New York: Oxford University Press

[4] Fraley D.S et al, Impact of acute renal failure on mortality in end-stage liver disease with or without transplant. Kideny Int, 1998. **54**:518-524.

[5] Granto E, Elinav H, Kohen R. Markers of oxidative stress in cyclosporine-treated and tacrolimus-treated children after liver transplantation. Liver Transpl. 2002 May; 8 (5): 469-75

[6] Hall, Daniel B. (2000). *Zero-Inflated Poisson and Binomial Regression with Random Effects: A Case Study*. Biometrics **56**, 1030-1039.

[7] Kellum, John A. et al. *RIFLE criteria for acute kidney injury are associated with hospital mortality in critically ill patients: a cohort analysis*. Critical Care 2006, **10**:73

[8] Lam, K.F. ; Xue, Hongqi Xue ; Cheung, Yin Bun (2006). *Semiparametric Analysis of Zero-Inflated Count Data. Biometrics*. (in review)

[9] Lambert, D. (1992). *Zero-inflated Poisson regression, with an application to defects in manufacturing.* Technometrics **34**, 1-14.

[10] Liang, Kung-Yee; Zeger, Scott L. (1986). *Longitudinal data analysis using generalized linear models*. Biometrika. **73**: 13-22.

[11] Lipsitz, Stuart R.; Kim, Kyungmann. (1994). *Analysis of Repeated Categroical Data Using Generalized Estimating Equations*. Statist. Medic. **13**: 1149-1163.

[12] Nelder, J. A. ;Wedderburn R. W. *Generalized linear models*. Journal of the Royal Statistical Society, Series A (General) **135** (3) (1972) 370—384

[13] Tatsuka, E. et al. *Influence of high donor serum sodium levels on early postoperative graft function in human liver transplantation: effect of correction of donor hypernatremia*. Liver Transpl Surg. 1999 Sept;5(5):421-8.

[14] Tonnesen A.S., Clinical pharmacology and use of diuretics. Review Courses in Anesthesiology, S.G. Hershy. B.J Bamforth and H. S Zauder. 1983, Philadelphia: J.B. Lippincott Co. 217-226

[15] Verbeke, Geert ; Molenberghs, Geert. (2005). Models for Discrete Longitudinal Data. New York: Springer.