International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 6 (Special Issue), 2007, pp 75-88

Some Nonparametric and Regression Model Strategies for Counts of Primary Events from Randomized Studies

Gary G. Koch
Dept. of Biostatistics
University of North Carolina
Chapel Hill, NC 27599-7420
U.S.A

Maura E. Stokes
SAS Institute Inc.,
SAS Campus Drive
Cary, NC 27513-2414
U.S.A

[Received September 5, 2005; Revised June 25, 2007; Accepted July 7, 2007]

Abstract

This paper discusses some nonparametric and regression model strategies for analyzing randomized studies where patients can have one or more occurrences of an unfavorable (or favorable) event. Principal attention is given to methods for counts of a recurrent event during inter-visit intervals which comprise an entire follow-up period. A major issue for confirmatory studies of this type is how to rank the patterns of events. A second issue is the management of missing data, particularly when patients can withdraw prematurely because of lack of efficacy or a terminating event such as death for which no treatment effects are expected. Both nonparametric methods and regression models fitted with generalized estimating equations (GEE) are useful for comparisons between treatments for the extent of one or more events for a primary response variable. Results from the described methods for counts of a recurrent event are illustrated for two confirmatory randomized clinical trials for comparing two treatments with respect to skeletal complications in patients with metastatic bone disease.

Keywords and Phrases: Logistic Regression for Repeated Events, Poisson Regression, Generalized Estimating Equations, Wilcoxon Rank Sum Tests for Event Rates.

AMS Classification: 62P10, 62J12.

1 Introduction

Many confirmatory randomized clinical trials and epidemiologic studies have a primary response variable for which each patient can have one or more occurrences of an unfavorable (or favorable) event. In most of these clinical trials, patients receive treatments to reduce the frequency of occurrences of a severe event (e.g. exacerbations of a respiratory disorder) during an entire follow-up period or during inter-visit intervals. The best possible outcome for these patients is 0 occurrences (or complete prevention of events) during the entire follow-up period, and many occurrences throughout the follow-up period is a very unfavorable outcome. Between these extremes, some not overly severe events during the early part of the follow-up period and no events in the latter part can be a favorable outcome. Some examples of studies with data for one or more occurrences of a recurrent event during an entire follow-up period or during inter-visit intervals are as follows:

1. lower respiratory illnesses (LRI) in children during the first year of life
   (LaVange et al [1994])

2. unscheduled medical visits for patients receiving treatment for asthma
   (Malmstrom et al [1999])

3. skeletal complications needing medical interventions for patients with metastatic bone disease
   (Moecks et al [2004])

There are several issues which require attention in plans for analyses of confirmatory randomized clinical trials where each patient can have one or more events for a primary response variable. Since the principal objective of a confirmatory study usually is the evaluation of whether patients with a test treatment have better outcomes than those with a control treatment, the most central consideration for analysis is the criterion for ranking the patterns of patient outcomes. In this regard, 0 occurrences during the entire follow-up period is the best outcome for a patient and many occurrences throughout the follow-up period is a very unfavorable outcome, but how to rank patterns of patient outcomes between these extremes can be difficult. Two issues which such a ranking needs to address are the management of incomplete follow-up (that is, missing data) for patients, and the management of multiple events during sub-intervals of time, particularly when their frequency can be outlyingly large. A noteworthy advantage of a justifiable ranking is that its relationship to test versus control treatments can be evaluated with nonparametric statistical methods without any formal assumptions about the underlying data structure. Complementary analyses through appropriate regression models are often additionally possible when the criterion for ranking is a count of events or an event rate per unit time.

The criteria for ranking the patterns of outcomes for patients include

1. time to first event
2. number of events
3. rate per unit of time for events

The extent of incomplete follow-up can adversely influence the previously stated criteria for ranking patients by causing patients to have a possibly smaller number of observed events than would have occurred if patients had completed the entire follow-up period. For this reason, a major question for patients with incomplete follow-up is whether the withdrawal from follow-up suggests lack of efficacy and thereby a higher rate of events after withdrawal from follow-up. A related question is concerned with how the duration of follow-up affects the interpretation of 0 events, since longer follow-up increases the strength of evidence for 0 events and vice versa. Another issue is how to have a ranking manage follow-up terminating events such as deaths when the intent of treatment is the reduction of morbidity events with no expected effects on mortality. A ranking that manages deaths as the worst outcome has the limitations of not accounting for the intent of treatment and of making the basis for ranking a mixture of mortality and morbidity. Conversely, a ranking that ignores deaths does not fully account for the extent to which the follow-up experience of a patient is unfavorable.

A straightforward way to manage incomplete follow-up for patients is to use a specified principle to assign a number of events to the patient for the time from withdrawal of follow-up to the end of the entire follow-up period. This number is then added to the observed number of events for the patient during their actual follow-up so as to produce a projected total number of events for the entire follow-up period for the patient. This type of projected total can then be used as the basis for ranking patients according to the extent of their occurrences of events. Several potential principles for assigning a number of events for the time from withdrawal of follow-up to the end of the follow-up period are as follows:

1. if a patient no longer has the disorder being treated at the time of withdrawal, add 0 events to the number of events prior to withdrawal;
2. If withdrawal is unrelated to the occurrence of events, add $y(T-t)/t$ to the number of events y prior to withdrawal where t is the duration of follow-up and T is the planned duration of the entire follow-up period;
3. If withdrawal suggests lack of efficacy and the overall rate of events is low, add the maximum of $(1, y(T-t)/t)$ to y.
4. If withdrawal is comparable to the worst possible outcome, then manage the patient as having the worst possible rank for the entire follow-up period

5. If a patient has 0 events at withdrawal, then their management should provide better ranks for longer duration of follow-up

With respect to these principles, there should be recognition that more stringent methods for managing withdrawals may overly favor test treatment when withdrawal tends to occur more frequently and earlier for a control group than a group with test treatment; and so a neutral method like (2) may be conservative. An additional consideration is that the use of more stringent methods for managing withdrawals with test treatment than for controls (e.g., (3) for test and (2) for control) can shed light on the robustness of results from comparisons between treatments.

When the possibly multiple events for a patient occur in a clearly distinct and not overly frequent manner, then their total number can provide a useful basis for ranking patients. However, in some situations, an outlyingly large (or impossible to enumerate) frequency of events can occur in a sub-interval of time as a consequence of a single underlying event, and such outliers can adversely influence the ranking of a patient by distorting their total number of events. One way to address this issue is to partition the entire follow-up period into a set of mutually exclusive intervals. Then each time interval is classified according to whether it has at least one event or according to its most severe event (with no event being the most favorable outcome and with one possibility for a severe event being an outlying frequency of events). The ranking of patients is then based on the number of time intervals with events (or severe events) rather than the number of events with outliers being avoided.

Given that a justifiable ranking is produced from appropriate ways of addressing incomplete follow-up for patients and multiple events during sub-intervals of time, nonparametric statistical methods can be used to compare the test and control treatments. In this regard, an important property of nonparametric statistical tests for treatment comparisons in a randomized clinical trial is that no formal assumptions are required (see Koch *et al* [1998]), mainly because the probabilistic structure for the test is a consequence of randomization in the study design. Well known nonparametric statistical tests for a ranking include the Wilcoxon rank sum test and its extended Mantel-Haenszel (or Van Elteren) extension to studies with stratified randomization (see Stokes *et al* [2000, Chapter 4]). Additional extensions to enable adjustment for continuous covariables or more covariables than stratification can accommodate are discussed in Koch *et al* [1998] and Tangen and Koch [2000].

When the criterion for ranking patterns of patient outcomes is a count of events for the follow-up period, an event rate per unit time, or such quantities during successive time intervals, methods based on generalized estimating equations (GEE) can be used to fit regression models to describe relationships to treatments and other explanatory variables (see Diggle et al [1994], and Royall [1986]). For such analyses, the sample size needs to be sufficiently large (for example, $\geq$ 100 patients) to support approximately normal distributions for estimates of parameters (although robustness for such approximations can be produced for treatment comparisons by using re-randomization methods such as those discussed by Westfall et al [1999]). GEE methods additionally

provide the empirical sandwich estimate for the covariance matrix of the estimated parameters with the property of consistency for this purpose having robust applicability regardless of the correctness of any specified working variance (or correlation) structure for the counts of events (with such robustness encompassing overdispersion as well). Missing data can be maintained as missing if considered completely at random, or can be managed with assigned (or extrapolated) values according to specified principles (for example, worst value, previously observed value, etc.). When missing data are maintained as missing, the underlying assumption is that the model adequately predicts both the quantities that are observed and those that are missing. The main consideration for justifying a specified principle for assigning values to missing data is that any bias in it is in favor of the control treatment and against the test treatment. An advantage of GEE methods relative to nonparametric statistical tests is that, in addition to p-values, they provide estimates of treatment differences and corresponding confidence intervals. As stated previously, an advantage of nonparametric methods is their much weaker structure for assumptions.

Additional references which discuss other methods for the statistical analysis of studies with one or more occurrences of primary events include Anderson et al [1993], Mathe and Chevret [1999], and Therneau and Hamilton [1997].

2 Metastatic Bone Disease Example

The previously described considerations for the statistical analysis of one or more primary events are well illustrated by two studies to compare test and control treatments for the reduction of the extent of skeletal complications for patients with metastatic bone disease (Moecks et al [2004]). Each study had 8 visits at 3 month intervals over a two-year follow-up period. The sample sizes for Study 1 were $n_C = 143$ for control and $n_T = 154$ for test, and the sample sizes for Study 2 were $n_C = 158$ for control and $n_T = 154$ for test. Events were based on a composite endpoint for medical interventions against bone pain or incident fractures.

For avoidance of excess counting of multiple events as a consequence of a single underlying event, each of the inter-visit intervals was classified as having at least one event or not. This data structure in terms of classifications for inter-visit intervals also accounted for diagnostic procedures such as X-rays only having planned use at the end of inter-visit intervals. Deaths were not a direct component of the composite endpoint (that is, they were managed as a random cause of incomplete follow-up (or censoring)) because the test treatment was not expected to have any effect on mortality. A substantial number of patients in each study withdrew prematurely because of death, signs of progression of the disease, or other reasons. For Study 1, 38% for control and 41% for test did not complete at least 6 inter-visit intervals; and for Study 2, 59% for control and 47% for test did not complete at least 6 inter-visit intervals.

2.1 Treatment Comparisons with Wilcoxon Rank Sum Statistics

The criteria for ranking the patterns of outcomes for patients in the studies concerning skeletal complications were as follows:

1. the number of inter-visit intervals with at least one event (which assumes no events after withdrawal);
2. the rate (y/t) for the number of intervals with at least one event y relative to the number of intervals prior to withdrawal t
3. the "smoothed rate" based on $(y + 0.5)/(t + 1)$.

Use of the rate (y/t) involves the assumption that the event rate after withdrawal is the same as that prior to withdrawal. It also manages $(0/t) = 0$ as similarly informative regardless of t. The smoothed rate $(y + 0.5)/(t + 1)$ accounts for t in a manner similar to (y/t) but manages $y = 0$ as less informative when t is smaller.

The results from Wilcoxon rank sum statistics for comparisons between test and control for the criteria for ranking patterns of outcomes of patients in the studies concerning skeletal complications are shown in Table 1. For Study 1, the Wilcoxon rank sum p-values were less than 0.05 for all three ranking criteria with the result for the "smoothed rate" being somewhat more conservative through partly penalizing the test treatment for its slightly higher withdrawal rate. Conversely, the p-value for the "smoothed rate" was the strongest result for Study 2 by penalizing the placebo group for its partly higher withdrawal rate. This higher withdrawal rate for placebo needs some type of management through an event rate because its bias undermined the extent to which the simple count of intervals with at least one event could detect a significant difference between the test and control treatments.

Table 1. Results from Wilcoxon Rank Sum Statistics for Comparisons of Ranking Criteria between Test and Control

		Control		Test		
Study	**Criterion**	**Mean**	**Std Dev**	**Mean**	**Std Dev**	**Wilcoxon p**
1	y	1.27	1.38	0.79	1.10	0.002
1	(y/t)	0.25	0.28	0.16	0.25	0.002
1	$(y + 0.5)/(t + 1)$	0.28	0.22	0.23	0.19	0.017
2	y	1.15	1.29	0.94	1.22	0.077
2	(y/t)	0.28	0.31	0.20	0.28	0.018
2	$(y + 0.5)/(t + 1)$	0.33	0.22	0.27	0.20	0.005

2.2 Treatment Comparisons with GEE Methods

Methods based on generalized estimating equations (GEE) were used to fit logistic regression models to the probabilities of occurrence or not of at least one event dur-

ing the respective inter-visit intervals for each of the two studies concerning skeletal complications in patients with metastatic bone disease. For this purpose, SAS PROC GENMOD was used for independent subjects with specifications for one record per subject with a binomial distribution, logit link, a model which only included treatments for events/trials with y as events and t as trials. A repeated statement with subject corresponding to the unique identification numbers of the respective patients was invoked with an independence working correlation structure to produce the empirical sandwich estimate for the covariance matrix of the estimated model parameters. This method of covariance matrix estimation was used because of its robustness to overdispersion to the working variance of binomial distributions; see Appendix. For situations where events are enumerated throughout the continuous duration of the follow-up period (rather than the classification of inter-visit intervals as having at least one event or not), analogous methods involving Poisson regression are applicable (see Stokes et al [2000, Chapter 15.14] and Koch and Stokes [2004]).

The GEE methods had four different specifications for events/trials to address the varying numbers of periods (or intervals) at risk t of patients and the role of zero events. These specifications were as follows:

1. y as the actual number of events and t as the actual number of periods for follow-up

2. $(y + 0.5)$ as the number of events to manage 0 events and $(t + 1)$ as the actual number of periods for follow-up

3. modification of y to $[y + y(8 - t)/t] = (8y/t)$ for the projected number of events for 8 periods from the (y/t) rate for events per period and 8 as the number of periods

4. modification of y to $[(y + 0.5) + (y + 0.5)(8 - t)/(t + 1)] = 9(y + 0.5)/(t + 1)$ for the projected number of events for 9 periods from the $(y + 0.5)/(t + 1)$ rate for events per period and 9 as the number of periods

With specifications 3 and 4, each patient has equal weight in the estimation of model parameters whereas patients have weights proportional to t (or $(t + 1)$) with specifications 1 (or 2). The role of specifications 2 and 4 is to manage 0 events as more informative when t is larger. Results from methods for generalized estimating equations to fit logistic regression models to the probabilities of occurrence or not of at least one event during inter-visit intervals of studies concerning skeletal complications in patients with metastatic bone disease are shown in Table 2.

Table 2. Results from Methods for Generalized Estimating Equations from Alternative Specifications for Events and Periods

Study	Criterion	Periods	Estimate	95% Confidence Interval	Score p
1	y	t	0.60	(0.42, 0.84)	0.004
1	$(y+0.5)$	$(t+1)$	0.70	(0.55, 0.90)	0.006
1	$(8y/t)$	8	0.59	(0.40. 0.87)	0.007
1	$9(y+0.5)/(t+1)$	9	0.74	(0.58, 0.94)	0.017
2	y	t	0.68	(0.49, 0.94)	0.020
2	$(y+0.5)$	$(t+1)$	0.74	(0.59, 0.94)	0.014
2	$(8y/t)$	8	0.65	(0.46, 0.93)	0.020
2	$9(y+0.5)/(t+1)$	9	0.73	(0.59, 0.91)	0.006

All of the p-values for the comparisons between test and control are less than 0.05. Those from specifications 3 and 4 agree well with their counterparts from Wilcoxon rank sum tests in Table 1 since they similarly manage patients as having equal weights. The results from specifications 2 and 4 are somewhat weaker than those from specifications 1 and 3 for Study 1 (where the withdrawal rate for test treatment is higher than control) and somewhat stronger for Study 2 (where the withdrawal rate is higher for control). As noted previously, GEE methods provide estimates of odds ratios for the extent of lower probabilities of at least one event during inter-visit intervals for test versus control and corresponding confidence intervals in addition to p-values.

2.3 Discussion for GEE Methods

In this application, there is interest in analyzing the total number y of inter-visit intervals with at least one event (relative to the total number of intervals at risk) with the underlying longitudinal data structure ignored, particularly because information for the actual inter-visit intervals with events is not accurately available. Thus, the effects of explanatory variables at the patient level on the number of inter-visit intervals with events per interval at risk for the respective subjects is evaluated by using logistic regression models. Estimates for the parameters in these models are obtained by solving the maximum likelihood equations that correspond to the numbers of inter-visit intervals with events having independent binomial distributions (that is, by applying logistic regression as discussed in Stokes et al [2000, Chapter 8]), although no assumptions concerning underlying binomial distributions are involved. Through methods for generalized estimating equations (GEE), the estimated parameters have approximately normal distributions when the number of subjects is sufficiently large (e.g.,$\geq$ 100), and their covariance matrix is consistently estimated by the empirical sandwich estimate from GEE with robustness to the correctness of the working variance of binomial distributions (that is, overdispersion). However, the validity of the

previously stated results from GEE methods requires that the model for the numbers of inter-visit intervals with events per interval at risk is correct. For situations where the total number of inter-visit intervals with an event is the focus of analysis, the correctness of the model specification requires that all inter-visit intervals at risk are comparably affected by subject level explanatory variables (that is, no time × explanatory variable interaction) and that any time dependent explanatory variables have no effects. Also, if subjects have varying numbers of inter-visit intervals at risk, then such time intervals themselves must have no effects. Finally, as stated previously missing person time in this situation is considered as missing completely at random (or as adequately predicted from the data for observed person time), although it can alternatively be managed with assigned values according to specified principles (e.g., worst value, previously observed value, etc.).

The relationship between the probability of at least one event during a particular inter-visit interval and explanatory variables at both the level of the patient and at the time dependent interval level can be analyzed with GEE methods for repeated measures logistic regression. The respective patients are the primary sampling units (or subjects) and the respective intervals at risk are the observational units for analysis. Each interval has yes or no for at least one event as the response variable and the corresponding set of explanatory variables for treatment, interval, and treatment × interval interaction. The usual specifications for the estimation of model parameters with GEE methods are binomial distribution, logit link, a repeated statement with respect to patient identification numbers, and an exchangeable (or unstructured) working correlation structure. In this analysis, all intervals have equal (unit) weight, and so patients with more intervals for exposure have more weight in the determination of the estimated parameters. For the example in this paper, the underlying longitudinal data for the respective inter-visit intervals of the patients are not available, and so results from such longitudinal analyses are not presented.

3 Concluding Comments

Both nonparametric methods and regression models fitted with GEE methods can provide useful analyses of data from studies with one or more occurrences of one or more types of primary events for independent subjects. The value of such analyses is better when missing data are less extensive, have a known cause (e.g., death, treatment failure), or have a random nature. Large sample sizes make normal approximations through central limit theory more applicable to inferential results such as confidence intervals and p-values from statistical comparisons. For nonparametric methods, rerandomization methods can provide essentially exact p-values.

Acknowledgements

This paper is based on presentations at the Drug Information Association Meeting in Tokyo, Japan in August, 2002 and at the SAS User's Group International 29th Annual Conference in Montreal, Canada in May 2004.

References

1. Anderson, P.K., Borgan, O., Gill, R.D., Keiding, N. [1993]. *Statistical Methods Based on Counting Processes*. New York: Springer Verlag.

2. Diggle, P.J., Liang, K.Y., and Zeger, S.L. [1994]. *Analysis of Longitudinal Data*. Oxford: Oxford University Press.

3. Koch, G.G., Tangen, C.M., Jung, J.W., and Amara, I.A. [1998]. Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. *Statistics in Medicine*, 17:1863-1892.

4. Koch, G.G., and Stokes, M.E. [2004]. Paper 212-29: Some statistical strategies for analyzing confirmatory studies involving one or more occurrences of primary events. Proceedings of the SAS User's Group International 29^th Annual Conference.

5. LaVange, L.M., Keyes, L.L., Koch, G.G., and Margolis, P.A. [1994]. Application of sample survey methods for modeling ratios to incidence densities. *Statistics in Medicine*, 13:343-355.

6. Malmstrom, K., Rodriguez-Gomez, G., Guerra, J. *et al.* [1999]. Oral montelukast, inhaled beclomethasone, and placebo for chronic asthma: a randomized controlled trial. *Annals of Internal Medicine*, 130:487-495.

7. Mathe, C., Chevret, S. [1999]. Estimation of the treatment effect in a clinical trial when recurrent events define the endpoint. *Statistics in Medicine*, 18:1821-1829.

8. Moecks, J., Koch, G.G., Scott, M., Budde, M., Givens, S., Koehler, W., Knoessl, M. [2004]. Measures of morbidity in clinical studies with recurrent skeletal complications. *Journal of Biopharmaceutical Statistics*, 14(2):415-437.

9. Royall, R.M. [1986]. Model robust confidence intervals using maximum likelihood estimators. *International Statistical Review*, 54(2):221-226.

10. Stokes, M.E., Davis, C.S., and Koch, G.G. [2000]. *Categorical Data Analysis Using the SAS System*, 2nd ed. SAS Institute, Inc., Cary, N.C.

11. Tangen, C.M. and Koch, G.G. [2000]. Non-parametric covariance methods for incidence density analysis of time-to-event data from a randomized clinical trial and their complementary roles to proportional hazards regression. *Statistics in Medicine*, 19:1039-1058.

12. Therneau, T.M. and Hamilton, S.A. [1997]. RhDNase as an example of recurrent event analysis. *Statistics in Medicine*, 16:2029-2047.

13. Westfall, P.H., Tobias, R.D., Rom, D., Wolfinger, R.D., and Hochberg, Y. [1999]. Multiple comparisons and multiple tests using the SAS System. SAS Institute, Inc., Cary, N.C.

Appendix. Properties of estimated Parameters from Logistic Regression

Let y_j denote the number of events in T_j trials for patient j where $j = 1, 2, ..., n$. The y_j are assumed to be independent with expected values $E\{y_j\} = \mu_j = T_j\pi_j$ (and with all higher moments being finite), and the variation of the π_j is assumed to be well described by the logistic regression model $\pi_j = \exp(\boldsymbol{x_j'\beta})/\{1 + \exp(\boldsymbol{x_j'\beta})\}$ where $\boldsymbol{x_j'}$ is the j^{th} row of a specified full rank $(n \times t)$ matrix $\boldsymbol{X}$ and $\boldsymbol{\beta}$ is a $(t \times 1)$ vector of unknown parameters. Thus, if $\boldsymbol{y} = (y_1, ..., y_n)'$, $\boldsymbol{\mu} = (\mu_1, ..., \mu_n)'$, $\boldsymbol{\pi} = (\pi_1, ..., \pi_n)'$, and $\boldsymbol{T} = (T_1, ..., T_n)'$, then

$$E\{\boldsymbol{y}\} = \boldsymbol{\mu} = \boldsymbol{D_T\pi} = \boldsymbol{D_T D_\zeta^{-1}}\,\mathbf{exp(\boldsymbol{X\beta})}$$

where $\boldsymbol{\zeta} = \{\mathbf{1} + \mathbf{exp(\boldsymbol{X\beta})}\}$, and $\boldsymbol{D_a}$ is a diagonal matrix with respective diagonal elements from $\boldsymbol{a}$, $\mathbf{exp(\boldsymbol{a})}$ is the vector of exponentiated values of $\boldsymbol{a}$, and $\mathbf{1}$ is an $(n \times 1)$ vector of 1's.

If the y_j have independent binomial distributions relative to T_j as the number of trials (although this is not a necessary assumption), then the likelihood function would be

$$L(\boldsymbol{\pi}) = \prod_{j=1}^{n}\{T_j!/y_j!(T_j - y_j)!\}\pi_j^{y_j}(1 - \pi_j)^{(T_j - y_j)}.$$

Under the logistic regression model $\boldsymbol{\mu} = \boldsymbol{D_T D_\zeta^{-1}}\mathbf{exp(\boldsymbol{X\beta})}$, the corresponding log-likelihood and its derivatives with respect to $\boldsymbol{\beta}$ are as follows:

$$Log_e\{L(\boldsymbol{\beta})\} = \sum_{j=1}^{n}[y_j\boldsymbol{x_j'\beta} - T_j\{log_e(1 + \exp(\boldsymbol{x_j'\beta}))\} - \{log_e(T_j!/y_j!(T_j - y_j)!)\}]$$

$$\frac{\partial Log_e\{L(\boldsymbol{\beta})\}}{\partial \boldsymbol{\beta}'} = \sum_{j=1}^{n}[y_j \boldsymbol{x}_j' - \frac{T_j \exp(\boldsymbol{x}_j'\boldsymbol{\beta})\boldsymbol{x}_j'}{(1+\exp(\boldsymbol{x}_j'\boldsymbol{\beta}))}]$$

$$= \sum_{j=1}^{n}[y_j \boldsymbol{x}_j' - \mu_j(\boldsymbol{\beta})\boldsymbol{x}_j']$$

$$= \{\boldsymbol{y}'\boldsymbol{X} - [\boldsymbol{\mu}(\boldsymbol{\beta})]'\boldsymbol{X}\}$$

where $\boldsymbol{\mu}(\boldsymbol{\beta}) = \boldsymbol{D_T D_\zeta^{-1}} \exp(\boldsymbol{X\beta})$. On this basis, the maximum likelihood estimates $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ and $\hat{\boldsymbol{\mu}}$ for $\boldsymbol{\mu}$ satisfy the non-linear equations

$$\boldsymbol{X'y} = \boldsymbol{X'}[\boldsymbol{\mu}(\hat{\boldsymbol{\beta}})] = \boldsymbol{X'D_T D_{\zeta(\hat{\beta})}^{-1}} \exp(\boldsymbol{X}\hat{\boldsymbol{\beta}})$$

for which the solution to obtain $\hat{\boldsymbol{\beta}}$ requires iterative computing methods. These equations are typically called logistic regression estimating equations, and their solution $\hat{\boldsymbol{\beta}}$ is typically called the logistic regression vector of parameter estimates. However, the use of $\hat{\boldsymbol{\beta}}$ to estimate $\boldsymbol{\beta}$ does not require the y_j to have independent binomial distributions, although the applicability of independent binomial distributions enables $\hat{\boldsymbol{\beta}}$ to have optimal precision.

The behavior of $\hat{\boldsymbol{\beta}}$ is characterized in large samples by its linear Taylor series about $\boldsymbol{\beta}$ regardless of whether the y_j have binomial distributions. This linear Taylor series has the structure

$$\hat{\beta}_{TS}(\boldsymbol{y}) = \hat{\boldsymbol{\beta}}(\boldsymbol{\pi}) + [\frac{\partial \hat{\boldsymbol{\beta}}}{\partial \boldsymbol{y}'}|_{\boldsymbol{y}=\boldsymbol{D_T\pi}}](\boldsymbol{y} - \boldsymbol{D_T\pi})$$

where $[\frac{\partial \hat{\boldsymbol{\beta}}}{\partial \boldsymbol{y}'}]$ is determined from the logistic regression estimating equations; that is, $\frac{\partial}{\partial \boldsymbol{y}'}[\boldsymbol{X'y} = \boldsymbol{X'D_T D_{\zeta(\hat{\beta})}^{-1}} \exp(\boldsymbol{X}\hat{\boldsymbol{\beta}})]$ yields $\boldsymbol{X'} = \boldsymbol{X'D_T D_{\hat{\pi}} D_{(1-\hat{\pi})} X} \frac{\partial \hat{\boldsymbol{\beta}}}{\partial \boldsymbol{y}'}$, where $\hat{\boldsymbol{\pi}} = \boldsymbol{D_{\zeta(\hat{\beta})}^{-1}} \boldsymbol{exp}(\boldsymbol{X}\hat{\boldsymbol{\beta}})$ and so $[\frac{\partial \hat{\boldsymbol{\beta}}}{\partial \boldsymbol{y}'}|_{\boldsymbol{y}=\boldsymbol{D_T\pi}}] = (\boldsymbol{X'D_\nu X})^{-1}\boldsymbol{X'}$ with $\boldsymbol{\nu} = [T_1\pi_1(1-\pi_1), ..., T_n\pi_n(1-\pi_n)]'$. It follows that

$$\hat{\beta}_{TS}(\boldsymbol{y}) = \boldsymbol{\beta} + (\boldsymbol{X'D_\nu X})^{-1}\boldsymbol{X'}(\boldsymbol{y} - \boldsymbol{D}_T\boldsymbol{\pi}).$$

Since $\hat{\boldsymbol{\beta}}$ behaves like its linear Taylor series counterpart $\hat{\beta}_{TS}$ when sample sizes are large, $\hat{\boldsymbol{\beta}}$ approximately has the multivariate normal distribution with expected value vector $\boldsymbol{\beta}$ and covariance matrix

$$\boldsymbol{Var}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X'D_\nu X})^{-1}\boldsymbol{X'Var(y)X(X'D_\nu X)}^{-1}$$

when sample sizes are large enough for $\boldsymbol{X'(y - D_T\pi)}$ to have an approximately multivariate normal distribution via Liapounov central limit theorems. These properties do not require the y_j to have independent binomial distributions; but they do require the y_j to be independent, and they require the model

$$E\{y_j\} = \mu_j = T_j\pi_j = T_j \exp(\boldsymbol{x'_j\beta})/[1 + \exp(\boldsymbol{x'_j\beta})]$$

to be a correct specification. They also require n to be sufficiently large to support approximate normality for $\boldsymbol{X}'(\boldsymbol{y - D_T\pi})$.

When the y_j have independent binomial distributions, $\boldsymbol{Var(y) = D_\nu}$ and $\boldsymbol{Var}(\hat{\boldsymbol{\beta}})$ simplifies to $(\boldsymbol{X'D_\nu X})^{-1}$ for which a consistent estimator is $V_{B\hat{\beta}} = (\boldsymbol{X'D_{\hat{\nu}}X})^{-1}$ where $\hat{\boldsymbol{\nu}} = (T_1\hat{\pi}_1(1 - \hat{\pi}_1), ..., T_n\hat{\pi}_n(1 - \hat{\pi}_n))'$. More generally, the y_j do not have binomial distributions, and so a robust estimator for $\boldsymbol{Var}(\hat{\boldsymbol{\beta}})$ is needed. For this purpose, the empirical sandwich estimator (as provided by methods for generalized estimating equations (GEE)) is applicable. This sandwich estimator for $\boldsymbol{Var}(\hat{\boldsymbol{\beta}})$ is given by

$$\boldsymbol{V_{G\hat{\beta}}} = (\boldsymbol{X'D_{\hat{\nu}}X})^{-1}[\boldsymbol{X'D_vX}](\boldsymbol{X'D_{\hat{\nu}}X})^{-1}$$

where $\boldsymbol{D_v}$ is a diagonal matrix for which the diagonal elements are the respective $v_j = (y_j - \hat{\mu}_j)^2$ where $\hat{\mu}_j = T_j\hat{\pi}_j$ with $\hat{\pi}_j = \exp(\boldsymbol{x'_j\hat{\beta}})/[1 + \exp(\boldsymbol{x'_j\hat{\beta}})]$. This estimator is robust for $\boldsymbol{Var}(\hat{\boldsymbol{\beta}})$, but it could be unsatisfactorily crude unless the sample size n is sufficiently large (e.g., n $\geq$ 100).

One can further note that $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{V_{G\hat{\beta}}}$ remain the same if the y_j and the T_j are multiplied by any constant C; that is,

$$\boldsymbol{X'}[C\boldsymbol{y}] = \boldsymbol{X'D}_{C\boldsymbol{T}}\boldsymbol{D^{-1}_{\zeta(\hat{\beta})}exp(X\hat{\beta})}$$

for $C\boldsymbol{y}$ relative to $C\boldsymbol{T}$ simplifies by cancellation of C to

$$\boldsymbol{X'y} = \boldsymbol{X'D_T}\boldsymbol{D^{-1}_{\zeta(\hat{\beta})}exp(X\hat{\beta})};$$

and for $\boldsymbol{V_{G\hat{\beta}}}$, multiplication of T_j by C causes multiplication of $\hat{\nu}_j$ by C and thereby multiplication of $(\boldsymbol{X'D_{\hat{\nu}}X})^{-1}$ by $(1/C)$, but corresponding multiplication of y_j and T_j by C causes multiplication of $v_j = (y_j - \hat{\mu}_j)^2$ by C^2 and thereby multiplication of $(\boldsymbol{X'D_vX})$ by C^2, and so the overall multiplier for $\boldsymbol{V_{G\hat{\beta}}}$ is $(1/C)(C^2)(1/C) = 1$ in correspondence to the invariance of $\boldsymbol{V_{G\hat{\beta}}}$ to the multiplication of y_j and T_j by any

constant C.

The principal requirement for $\boldsymbol{V_{G\hat{\beta}}}$ to be a robust estimator for $\boldsymbol{Var(\hat{\beta})}$ is correctness of the logistic regression model

$$E\{y_j\} = \mu_j = T_j\pi_j = T_j \exp(\boldsymbol{x_j'\beta})/[1 + \exp(\boldsymbol{x_j'\beta})].$$

However, this requirement has the underlying assumption that the explanatory variables do not have different effects during successive time intervals with numbers of events y_{ij} and numbers of trials T_{ij} for $i = 1, 2, ..., d_j$ such that $y_j = \sum_{i=1}^{dj} y_{ij}$ and $T_j = \sum_{i=1}^{dj} T_{ij}$ (that is, no time $\times$ explanatory variable interaction applies). Also, missing counts of events for missing time intervals or for time intervals after discontinuation of follow-up (that is, $d_j < i \leq d$) are assumed to be missing completely at random (that is, they are assumed to be compatible with the model for the observed counts). An additional assumption when patients have varying numbers of trials (because of discontinuation of a study prior to completion of a specified follow-up period) is that the rates of events for the successive time intervals are homogeneous for each of the patients so that $E\{Y_{ij}/T_{ij}\} = (\mu_{ij}/T_{ij}) = \pi_{ij} = \pi_j$ since it implies $\mu_{ij} = \pi_j T_{ij}$ and $E\{y_j\} = \mu_j = \sum_{i=1}^{dj} \pi_j T_{ij} = T_j\pi_j$. Clarification of the necessity of this assumption is provided by consideration of the structure $\pi_{ij} = \theta_i\pi_j$ (with θ_i as a multiplicative effect for the i^{th} time interval) as a simple departure since it implies $\mu_{ij} = \theta_i\pi_j T_{ij}$ and $E\{y_j\} = \mu_j = \sum_{i=1}^{dj} E\{y_{ij}\} = \sum_{i=1}^{dj} \theta_i\pi_j T_{ij} = \pi_j \sum_{i=1}^{dj} \theta_i T_{ij}$; but this structure implies that the model $\mu_j = T_j\pi_j$ is an incorrect specification unless all $\theta_i = 1$ (or all $T_{ij} = T_i$ and all $d_j = d$ so that $\mu_j = \pi_j \sum_{i=1}^{d} \theta_i T_i = \pi_j K$ with $K = \sum_{i=1}^{d} \theta_i T_i$ being the same for all subjects).

The previously described considerations concerning the π_{ij} can be addressed to some extent by using GEE methods to fit a repeated measures logistic regression model to the y_{ij} relative to the T_{ij}. Whether missing data is missing completely at random would still be an issue.