Paper 212-29

Some Statistical Strategies for Analyzing Confirmatory Studies Involving One Or More Occurrences Of Primary Events

Gary G. Koch, University of North Carolina, Chapel Hill, NC
Maura E. Stokes, SAS Institute Inc., Cary, NC

Abstract

This presentation discusses some statistical strategies for analyzing confirmatory 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 examples. One is a pair of confirmatory randomized clinical trials for comparing two treatments with respect to skeletal complications in patients with metastatic bone disease. The second is an epidemiologic study concerning the association of lower respiratory illnesses in children during the first year of life with passive smoking in the home and other explanatory variables.

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 (for example, exacerbations of a respiratory disorder) during an entire follow-up period or during inter-visit intervals. The best possible outcome for these patients is zero 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:

- lower respiratory illnesses (LRI) in children during the first year of life (LaVange et al [1994], Stokes et al [2000, Chapter 15.14])

- fractures in post-menopausal women receiving treatment for osteoporosis
  (Stokes et al [2000, Chapter 15.9])
- daily seizure events in epilepsy
  (Albert and Follman [2000])
- unscheduled medical visits for patients receiving treatment for asthma
  (Itzler et al [2000], Malmstrom et al [1999])
- 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, zero 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

- a dichotomy for occurrence of at least one event or not
- time to first event
- number of events
- weighted (by severity) number of events
- rate per unit of time for events
- weighted (by severity) average of rates per unit of time for events
- time from last event to end of planned follow-up period

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 zero events, since longer follow-up increases the strength of evidence for zero 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 zero 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 zero 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 (that is, (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 accomodate are discussed in Koch et al [1998].

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], Liang et al [1992], Royall [1986], and Zeger and Liang [1986]). For such analyses, the sample size needs to be sufficiently large (for example, ≥ 100 patients) to support approximately normal distributions for estimates of parameters. 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 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], Cook and Lawless [1997], Lawless and Nadeau [1995], Mathe and Chevret [1999], Stukel et al [1994], Thall [1988], Thall and Lachin [1988], and Therneau and Hamilton [1997].

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 [2000]). 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.

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 penalyzing the test treatment for its slightly higher withdrawal rate. Conversely, the p-value for the "smoothed rate" was the most strongly significant result for Study 2 by penalyzing 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

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 during the respective inter-visit intervals for each of the two studies concerning skeletal complications in patients with metastatic bone disease. For this purpose, the GENMOD procedure in SAS/STAT was used with specifications for binomial distribution, logit link, and 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 indentification numbers of the respective patients was used 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 2. The GEE methods had four different specifications for events/trials to address the varying follow-up times 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 zero 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 zero 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

For all of the respective specifications, 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 that 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 and corresponding confidence intervals in addition to p-values.

Lower Respiratory Disease Example

Methods for one or more primary events in a longitudinal data structure with time dependent covariables can be well illustrated with an epidemiologic study of lower respiratory illness (LRI) during the first year of life of children (LaVange et al [1994], Stokes et al [2000]). This study had assessments of 294 children for one or more episodes of lower respiratory illness (cough, wheezing, rattling in the chest) during consecutive two week intervals of risk for the first year of life. There were 6115 such two week intervals (with those two week intervals which had continuing LRI from the immediately preceding two week intervals for a child being excluded from consideration because they were not at risk for a new LRI). A total of 197 lower respiratory illnesses (LRI) in all were observed during 208.4 person years of risk where an interval with new LRI was managed as 1 week at risk and all other intervals were managed as 2 weeks at risk. Passive smoke exposure, socio-economic status (that is, education of parents) and crowding in the home were explanatory variables for the child; and age and season during the one year follow-up period were time dependent explanatory variables. Three types of statistical methods which have had illustrative application to the lower respiratory illness (LRI) data are as follows:

1. Poisson regression for total number of LRI per child relative to their total person weeks of exposure (see Stokes et al [2000]);

2. repeated measures logistic regression (via generalized estimating equations (GEE)) for LRI per two-week interval (see LaVange et al [1994]);

3. weighted least squares analysis of variance for correlated ratio estimates of incidence densities for LRI (see LaVange et al [1994]).

Poisson Regression for Total Number of Events

In some situations, there is interest in analyzing the total number of occurrences of an event such as LRI (relative to the total person weeks of exposure) with the underlying longitudinal data structure ignored, particularly when information for the times of occurrence of events is not accurately available. The effects of explanatory variables at the subject level on the numbers of events per unit of person time for the respective subjects can be evaluated by using loglinear regression models. Estimates for the parameters in these models are obtained by solving the maximum likelihood equations that correspond to the total numbers of events having independent Poisson distributions (that is, by applying Poisson regression as discussed in Koch et al [1986]), although no assumptions concerning underlying Poisson 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 (for example,$\geq$ 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 Poisson distributions (that is, overdispersion).

However, the validity of the previously stated results from GEE methods requires that the model for the numbers of events per unit of person time is correct. For situations where the total number of occurrences of an event is the focus of analysis, the correctness of the model specification requires that all units of person time are comparably affected by subject level explanatory variables (that is, no time $\times$ explanatory variable interaction) and that any time dependent explanatory variables have no effects. Also, if subjects have varying amounts of person time, then the 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 (that is, worst value, previously observed value, etc.).

A loglinear model that was analyzed for the total number of occurrences of LRI included explanatory variables for passive smoke exposure, crowding, socio-economic status (three categories for education of parents), race, and age group (three categories for mode during follow-up) at the level of the child. The GENMOD procedure was used to estimate the parameters in this model through specifications of Poisson distribution, log link, and log(time at risk in years) as the offset, although their role is to determine the estimating equations since no assumptions about the applicability of Poisson distributions is needed. Since over-dispersion was suggested (deviance value/d.f. = 1.48, Pearson value/d.f. = 1.80), GEE methods were used to obtain the empirical sandwich estimate for the covariance matrix of the estimated parameters in the model by putting subject identification numbers (that is, id) in the CLASS statement and specifying REPEATED SUBJECT=ID/TYPE=IND for independence as the working correlation matrix since there is only one record per subject.

The estimate of the log(incidence density ratio) for passive smoking (yes vs. no) was 0.431 with unadjusted standard error 0.165 ($p = 0.009$); the Pearson scaling adjusted standard error was 0.221 (relative to which $p = 0.052$), and the GEE standard error was 0.211 (relative to which $p = 0.041$); and the GEE score statistic yielded $p = 0.048$. The corresponding incidence density ratio was $exp(0.431) = 1.54$, with confidence interval (1.02, 2.32) from the GEE standard error. The advantages of the GEE standard errors relative to those from Pearson (or deviance scaling) are their versatility whereby each parameter estimate has its own adjustment of its standard error for over-dispersion and their robustness to the correctness of the specifications of working variances that accompany the estimating equations. See Appendix 1 for additional technical discussion.

Repeated Measures Logistic Regression

The relationship between the probability of an LRI during a particular two week interval and explanatory variables at both the level of the child and at the time dependent two week interval level was analyzed with GEE methods for repeated measures logistic regression. The respective children were the primary sampling units (or subjects) and two-week intervals with risk for LRI were the observational units for analysis. Each two-week interval had yes or no for LRI as the response variable and the corresponding set of explanatory variables for socio-economic status (SES), crowding in the home, age, season, and passive smoking (with age and season being time dependent).

The specifications for the estimation of model parameters with GEE methods were binomial distribution, logit link, a REPEATED statement with respect to subject identification numbers, and an exchangeable working correlation structure. In this analysis, all two week intervals had equal (unit) weight, and so children with more two-week intervals for exposure had more weight in the determination of the estimated parameters. The results from GEE methods are shown in Table 3. The effect of passive smoking corresponded to an odds ratio of exp(0.44) = 1.56 with 95% confidence interval (1.02, 2.38). These results are similar to their counterparts from Poisson regression for the total number of LRI in the previous section, but they have the advantage of being based on a model that accounts for the strong effects of age and season as time dependent explanatory variables.

Table 3. Results from GEE Methods for Longitudinal Study of Lower Respiratory Illness (LRI)

Parameter	Estimate	Standard Error	p-value
Intercept	-4.82	0.24	
Lower SES	0.32	0.26	0.210
Crowding	0.47	0.23	0.040
Age: 4-6 months	0.80	0.22	< 0.001
> 6 months	0.63	0.20	0.002
Fall/winter	0.44	0.14	0.002
Passive Smoking	0.44	0.22	0.040

Weighted Least Squares Methods for Correlated Ratio Estimates of Incidence Densities

As described in Appendix 3, an incidence density for the number of LRI per week of person time can be estimated as the ratio of the mean number of LRI per child divided by the mean number of person weeks per child. Such estimation is applicable to the respective conditions that correspond to a cross-classification of explanatory variables. For the longitudinal study of LRI, such estimates were determined for the cross-classification of (passive smoking x season x age group) together with a consistent estimate for the corresponding covariance matrix. Since the respective ratios were based on at least 100 children and 80% of them were based on ≥ 5 children with ≥ 1 LRI (with the others being based on ≥ 3 children with ≥ 1 LRI), the ratio estimates for the respective incidence densities approximately have a multivariate normal distribution. The results from the use of the CATMOD procedure to fit a linear model with additive effects for age, season, and passive smoking to the natural logarithms of the incidence densities are shown in Table 4.

The effect of passive smoking corresponded to an incidence density ratio $\exp(0.71) = 2.04$ with 95% confidence interval (1.40, 2.97); that is, the rate of LRI for children in homes with passive smoking was about twice the rate as for homes without passive smoking. However, this analysis has the limitation of not accounting for crowding in the home and socio-economic status because its sample size requirements to support approximately normal distributions imply that its scope can only address a small number of explanatory variables. Nevertheless, a noteworthy advantage of these methods is that they do not require any assumptions about the distributions of the numbers of LRI for the cross-classification of conditions which they address.

Table 4. Results from Weighted Least Squares Estimation of Parameters in Linear Model for Logarithms of Incidence Densities

Parameter	Estimate	Standard Error	p-value
Intercept	-1.19	0.23	
Age $>$ 6 months	0.56	0.21	0.006
Age 4-6 months	0.83	0.21	$<$ 0.001
Fall/Winter	0.47	0.14	0.001
Passive Smoking	0.71	0.19	$<$ 0.001

Nevertheless, a noteworthy advantage of these methods is that they do not require any assumptions about the distributions of the numbers of LRI for the cross-classification of conditions which they address.

The principal consideration which guides the use of weighted least squares analysis of variance for correlated ratio estimates of incidence densities is that its scope is limited to a small number of categorical factors for conditions and groups because large sample sizes are necessary to support approximately multivariate normal distributions for such estimates. Advantages of the methods are that no technical assumptions are necessary in the setting where the ratio estimates are inherently meaningful and that both numerators and denominators in the ratio estimates are managed as

random. The method operates so that subjects with larger amounts of person time have larger weights in the ratio estimates for the incidence densities. Also, an underlying assumption is that the same incidence densities apply to unobserved exposure times with missing event counts as to the corresponding observed ones (that is, missing information is missing completely at random). In this regard, the extent to which the occurrence of an event affects duration of overall follow-up or the rate of subsequent events (when duration of follow-up is not complete) can merit attention.

Extensions for Nonparametric Covariance Adjustment of Ratio Estimates from Randomized Studies

The methods described in Tangen and Koch [2000] enable correlated ratio estimates of incidence densities from a randomized clinical trial to have nonparametric covariance adjustment for both continuous and categorical baseline factors without an increase for the number of groups (or conditions) being evaluated. For example, with two groups, analysis is focused on the vector $(\mathbf{F}'_{\mathbf{h}}, \bar{\mathbf{Z}}'_{\mathbf{h}})'$ for the h^{th} group where $\mathbf{F_h}$ denotes the vector of natural logarithms of ratio estimates for incidence densities and $\bar{\mathbf{Z}}_{\mathbf{h}}$ denotes the vector of means for the baseline covariables which are to have adjustment. Weighted least squares is used to fit linear models which impose no differences among the groups for the $\bar{\mathbf{Z}}_{\mathbf{h}}$ on the basis of randomization in the study design. The resulting estimates $\mathbf{b_h}$ which correspond to the $\mathbf{F_h}$ have covariance adjustment in the sense of pertaining to a structure for which there is no difference between the groups for the means of the covariables. Thus, the result from the comparison of the $\mathbf{b_h}$ for the groups has covariance adjustment. The covariance matrix for the $(\mathbf{F}'_{\mathbf{h}}, \bar{\mathbf{Z}}'_{\mathbf{h}})'$ can be estimated more stably under the null hypothesis of no differences among treatments on the basis of the randomized design of the study so as to improve the applicability of chi-square approximations for tests to compare treatments and to enable the use of exact p-values in some sense. Also, adjustment for a set of strata is possible, missing information can be managed in sensitivity analyses through alternative conventions for enumerating events or specifying amounts of person time for unobserved time intervals, and the scope can be extended to encompass counts for events at two or more severities.

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. The value of such analyses is better when missing data are less extensive, have a known cause (that is, 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.

References

1. Albert, P.S., Follmann, D.A. [2000]. Modeling repeated count data subject to informative dropout. *Biometrics*, 56:667-677.

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

3. Cook, R.J., Lawless, J.F. [1997] Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine*, 16:911-924.

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

5. Itzler, R., Dasbach, E., Koch, G.G., Heyse, J., Cool, J., Davies, G. [2000]. Using synthetic GEE to estimate variability in pharmacoeconomic studies: the case of asthma. Presentation at the Seventh Annual Biopharmaceutical Applied Statistics Symposium (BASS), Hilton Head, SC.

6. Koch, G.G., Atkinson, S.S., and Stokes, M.E. [1986]. Poisson regression, in *Encyclopedia of Statistical Sciences*, 7, eds. Kotz, S. and Johnson, N.L., New York: John Wiley and Sons, Inc., 32-41.

7. 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.

8. 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.

9. Liang, K.Y., Zeger, S.L., and Qaqish, B.F. [1992]. Multivariate regression models for categorical data. *Journal of the Royal Statistical Society, Series B* 54:3-40.

10. Lawless, J.F., Nadeau, C. [1995]. Some simple robust methods for the analysis of recurrent events. *Technometrics*, 37:158-168.

11. 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.

12. 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.

13. 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. Submitted to *Journal of Biopharmaceutical Statistics*, in press.

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

15. 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.

16. Stukel, T.A., Glynn, R.J., Fisher, E.S., Sharp, S.M., Lu-Yao, G., Wennberg, J.E. [1994]. Standardized rates of recurrent outcomes. *Statistics in Medicine*, 13:1781-1791.

17. 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.

18. Thall, P.F. [1988]. Mixed Poisson likelihood regression models for longitudinal interval count data. *Biometrics*, 44:197-209.

19. Thall P.F., Lachin, J.M. [1988]. Analysis of recurrent events: non-parametric methods for random-interval count data. *Journal of the American Statistical Association*, 83:339-347.

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

21. Zeger, S.L., Liang, K.Y. [1986]. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42:121-130.

Appendix 1. Properties of Estimates Parameters from Poisson Regression

Let y_j denote a count of events for patient j during follow-up time T_j where $j = 1, 2, ..., n$. The y_j are assumed to be independent with expected values $E\{y_j\} = \mu_j = T_j\lambda_j$ (and with all higher moments being finite), and the variation of the λ_j is assumed to be well described by the loglinear model $\lambda_j = \exp(\mathbf{x_j'}\beta)$ where $\mathbf{x_j'}$ is the j^{th} row of a specified full rank $(n \times t)$ matrix $\mathbf{X}$. Thus, if $\mathbf{y} = (\mathbf{y_1}, ..., \mathbf{y_n})'$, and $\mu = (\mu_1, ..., \mu_n)'$, and $\mathbf{T} = (T_1, ..., T_n)'$, then $E\{\mathbf{y}\} = \mu = \mathbf{D_T}\exp(\mathbf{X}\beta)$ where $\mathbf{D_T}$ is a diagonal matrix with respective diagonal elements from $\mathbf{T}$ and $\exp(\mathbf{a})$ denotes the vector of exponentiated values of $\mathbf{a}$.

If the y_j have independent Poisson distributions (although this is not a necessary assumption), then the likelihood function would be

$$L(\mu) = \prod_{j=1}^{n} \mu_j^{y_j}(\exp(-\mu_j))/y_j!.$$

Under the loglinear model $\mu = \mathbf{D_T exp}(\mathbf{X}\beta)$, the corresponding log-likelihood is

$$\log\{L\{(\beta)\} = \sum_{j=1}^{n}[y_j\{\log T_j + \mathbf{x_j'}\beta\} - T_j\exp(\mathbf{x_j'}\beta) - \log(y_j!)].$$

Since

$$\partial\log\{L(\beta\}/\partial\beta' = \sum_{j=1}^{n}[y_j\mathbf{x_j'} - T_j\exp(\mathbf{x_j'}\beta)\mathbf{x_j'}] = \mathbf{y'X} - [\mu(\beta)]'X\}$$

where $\mu(\beta) = \mathbf{D_T}\exp(\mathbf{X}\beta)\}$, the maximum likelihood estimates $\hat{\beta}$ for β and $\hat{\mu}$ for μ satisfy the non-linear equations

$$\mathbf{X'y} = \mathbf{X'}[\hat{\mu}(\hat{\beta})] = \mathbf{X'D_T}\exp(\mathbf{X}\hat{\beta})$$

for which the solution to obtain $\hat{\beta}$ requires iterative computing methods. These equations are typically called Poisson regression estimating equations, and their solution $\hat{\beta}$ is typically called the Poisson regression vector of parameter estimates. However, the use of $\hat{\beta}$ to estimate β does not require the y_j to have independent Poisson distributions, although the applicability of independent Poisson distributions enables $\hat{\beta}$ to have optimal precision (see Royall [1986] and Diggle [1994]).

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

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

where $[\partial\hat{\beta}/\partial\mathbf{y}']$ is determined from the Poisson regression estimating equations, that is,

$$\frac{\partial}{\partial\mathbf{y}'}[\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{D_T}\exp(\mathbf{X}\hat{\beta})]$$

yields

$$\mathbf{X}' = \mathbf{X}'\mathbf{D_T}\mathbf{D}_{\exp(\mathbf{X}\hat{\beta})}\mathbf{X}\frac{\partial\hat{\beta}}{\partial\mathbf{y}'}$$

and so

$$[\frac{\partial\hat{\beta}}{\partial\mathbf{y}'}|_{\mathbf{y}=\mu}] = (\mathbf{X}'\mathbf{D}_\mu\mathbf{X})^{-\mathbf{1}}\mathbf{X}'$$

It follows that $\hat{\beta}_{TS}(\mathbf{y}) = \hat{\beta}(\mu) + (\mathbf{X}'\mathbf{D}_\mu\mathbf{X})^{-\mathbf{1}}\mathbf{X}'(\mathbf{y}-\mu) = \beta + (\mathbf{X}'\mathbf{D}_\mu\mathbf{X})^{-\mathbf{1}}\mathbf{X}'(\mathbf{y}-\mu)$. Since $\hat{\beta}$ behaves like its linear Taylor series counterpart $\hat{\beta}_{\mathbf{TS}}$ when sample sizes are large, $\hat{\beta}$ approximately has the multivariate normal distribution with expected value vector β and covariance matrix

$$\mathrm{Var}(\hat{\beta}) = (\mathbf{X}'\mathbf{D}_\mu\mathbf{X})^{-1}\mathbf{X}'\mathrm{Var}(y)X(\mathbf{X}'\mathbf{D}_\mu\mathbf{X})^{-1}$$

when sample sizes are large enough for $\mathbf{X}'(\mathbf{y}-\mu)$ to be approximately multivariate normally distributed via Liapounov central limit theorems. These properties do not require the y_j to have independentPoisson distributions; but they do require the y_j to be independent and they require the model $E\{y_j\} = \mu_j = T_j\exp(\mathbf{x}_\mathbf{j}'\beta)$ to be a correct specification. They also require n to be sufficiently large to support approximate multivariate normality for $\mathbf{X}'(\mathbf{y}-\mu)$.

When the y_j have independent Poisson distributions, $\mathrm{Var}(n) = \mathbf{D}_\mu$, and $Var(\hat{\beta})$ simplifies to $(\mathbf{X}'\mathbf{D}_\mu\mathbf{X})^{-1}$ for which a consistent estimator is $\mathbf{V}_{\mathbf{p}\hat{\beta}} = (\mathbf{X}'\mathbf{D}_{\hat{\mu}}\mathbf{X})^{-1}$ where $\hat{\mu} = \mathbf{D_T}\exp(\mathbf{X}\hat{\beta})$. More generally, the y_j do not have Poisson distributions, and so a robust estimator for $\mathrm{Var}(\hat{\beta})$ is needed. For this purpose, the empirical sandwich variance estimator (as provided by methods for generalized estimating equations (GEE)) is applicable. This sandwich estimator for $\mathrm{Var}(\hat{\beta})$ is

$$\mathbf{V}_{\mathbf{G}\hat{\beta}} = (\mathbf{X}'\mathbf{D}_{\hat{\mu}}\mathbf{X})^{-1}[\mathbf{X}'\mathbf{D}_\mathbf{v}\mathbf{X}](\mathbf{X}'\mathbf{D}_{\hat{\mu}}\mathbf{X})^{-1}$$

where $\mathbf{D_v}$ is a diagonal matrix for which the diagonal elements are the respective $v_j = (y_j - \hat{\mu}_j)^2$. This estimator is robust for $\mathrm{Var}(\hat{\beta})$, but it could be unsatisfactorily crude unless the sample size n is sufficiently large (that is, $n \geq 100$). The principal requirement for $\mathbf{V}_{\mathbf{G}\hat{\beta}}$ to be a robust estimator for $\mathrm{Var}(\hat{\beta})$ is correctness of the model $\mu = \mathbf{D_T}\exp(\mathbf{X}\beta)$. 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 amounts of follow-up time 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 incomplete observation of a time interval (that is, $T_{ij} \leq T_i$ where T_i is the length of the i^{th} interval where $i = 1, 2, ..., d$) 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 follow-up times (because of discontinuation of a study prior to completion of a specified follow-up period) is that the rates of events per unit of person time for the successive time intervals are homogeneous for each of the patients so that $E\{Y_{ij}/T_{ij}\} = (\mu_{ij}/T_{ij}) = \lambda_{ij} = \lambda_j$ since it implies $\mu_{ij} = \lambda_j T_{ij}$ and $E\{y_j\} = \mu_j = \sum_{i=1}^{dj} \lambda_j T_{ij} = T_j\lambda_j$. Clarification of the necessity of this assumption is provided by consideration of the structure $\lambda_{ij} = \theta_i\lambda_j$ (with θ_i as a multiplicative effect for the i^{th} time interval) as a simple departure since it implies $\mu_{ij} = \theta_i\lambda_j T_{ij}$ and $E\{y_j\} = \mu_j = \sum_{i=1}^{dj} E\{y_{ij}\} = \sum_{i=1}^{dj} \theta_i\lambda_j T_{ij} = \lambda_j \sum_{i=1}^{dj} \theta_i T_{ij}$; but this structure implies that the loglinear model $\mu_j = T_j \exp(\mathbf{x_j'}\beta)$ is an incorrect specification unless all $\theta_i = 1$ (or all $T_{ij} = T_i$ and all $d_j = d$ so that $\mu_j = \lambda_j \sum_{i=1}^{d} \theta_i T_i = \lambda_j C$ with $C = \sum_{i=1}^{d} \theta_i T_i$ being the same for all subjects).

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

Appendix 2. 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 $\pi_j = \exp(\mathbf{x_j'}\beta)/\{1 + \exp(\mathbf{x_j'}\beta)\}$ where $\mathbf{x_j'}$ is the j^{th} row of a specified full rank $(n \times t)$ matrix $\mathbf{X}$. Thus, if $y = (y_1, ..., y_n)'$, $\mu = (\mu_1, ..., \mu_n)'$, $\pi = (\pi_1, ..., \pi_n)'$, and $\mathbf{T} = (T_1, ..., T_n)'$, then

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

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

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

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

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

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

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

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

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

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

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

for which the solution to obtain $\hat{\beta}$ requires iterative computing methods. These equations are typically called logistic regression estimating equations, and their solution $\hat{\beta}$ is typically called the logistic regression vector of parameter estimates. However, the use of $\hat{\beta}$ to estimate β does not require the y_j to have independent binomial distributions, although the applicability of independent binomial distributions enables $\hat{\beta}$ to have optimal precision.
The behavior of $\hat{\beta}$ is characterized in large samples by its linear Taylor series about β regardless of whether the y_j have binomial distributions. This linear Taylor series has the structure

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

where $[\frac{\partial \hat{\beta}}{\partial \mathbf{y}'}]$ is determined from the logistic regression estimating equations; that is,

$\frac{\partial}{\partial \mathbf{y}'}[\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{D_T D_{\zeta(\hat{\beta})}^{-1}} \exp(\mathbf{X}\hat{\beta})]$ yields $\mathbf{X}' = \mathbf{X}'\mathbf{D_T D_{\hat{\pi}} D_{(1-\hat{\pi})}} \mathbf{X} \frac{\partial \hat{\beta}}{\partial \mathbf{y}'}$,

and so $[\frac{\partial \hat{\beta}}{\partial \mathbf{y}'}|_{\mathbf{y}=\mathbf{D_T}\pi}] = (\mathbf{X}'\mathbf{D}_\nu\mathbf{X})^{-1}\mathbf{X}'$ where $\hat{\pi} = \mathbf{D_{\zeta(\hat{\beta})}^{-1}} \mathbf{exp}(\mathbf{X}\hat{\beta})$ and

$\nu = [T_1\pi_1(1-\pi_1), ..., T_n\pi_n(1-\pi_n)]'$. It follows that

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

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

$$\mathrm{Var}(\hat{\beta}) = (\mathbf{X}'\mathbf{D}_\nu\mathbf{X})^{-1}\mathbf{X}'\mathbf{Var}(\mathbf{y})\mathbf{X}(\mathbf{X}'\mathbf{D}_\nu\mathbf{X})^{-1}$$

when sample sizes are large enough for $\mathbf{X}'(\mathbf{y} - \mathbf{D_T}\pi)$ 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

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

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

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

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

where $\mathbf{D_v}$ 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 $\hat{\pi}_j = \exp(\mathbf{x_j'}\hat{\beta})/[1 + \exp(\mathbf{x_j'}\hat{\beta})]$. This estimator is robust for $\text{Var}(\hat{\beta})$, but it could be unsatisfactorily crude unless the sample size n is sufficiently large (that is, n $\geq$ 100). The principal requirement for $\mathbf{V_{G\hat{\beta}}}$ to be a robust estimator for $\text{Var}(\hat{\beta})$ is correctness of the logistic regression model

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

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

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

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

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

and for $\mathbf{V_{G\hat{\beta}}}$, multiplication of T_j by C causes multiplication of $\hat{\nu}_j$ by C and thereby multiplication of $(\mathbf{X'D}_{\hat{\nu}}\mathbf{X})^{-1}$ 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 $(\mathbf{X'D_vX})$ by C^2, and so the overall multiplier

for $\mathbf{V}_{\mathbf{G}\hat{\beta}}$ is $(1/C)(C^2)(1/C) = 1$ in correspondence to the invariance of $\mathbf{V}_{\mathbf{G}\hat{\beta}}$ to the multiplication of y_j and T_j by any constant C.

Appendix 3. Weighted Least Squares Analysis for Ratio Estimates of Incidence Densities

Weighted least squares analysis of variance is applicable to ratio estimates of incidence densities (LaVange et al [1994], Tangen and Koch [2000]). For this method, let y_{ghk} = (number of events) for patient k in group h during time condition g applies; T_{ghk} = (amount of exposure time) for patient k in group h during time condition g applies; $y_{ghk} = T_{ghk} = 0$ when patient k in group h has no time for exposure to events during condition g; $g = 1, 2, ..., G$; $h = 1, 2, ..., H$; $k = 1, 2, ..., n_h$.

Let

$$\bar{\mathbf{y}}_h = \{\sum_{k=1}^{n_h}(y_{1hk}, ..., y_{Ghk})'/n_h\} = \{\sum_{k=1}^{n_h}\mathbf{y}_{hk}/n_h\} = (\bar{y}_{1h}, ..., \bar{y}_{Gh})'$$

and

$$\bar{\mathbf{T}}_h = \{\sum_{k=1}^{n_h}(T_{1hk}, ..., T_{Ghk})'/n_h\} = \{\sum_{k=1}^{n_h}\mathbf{T}_{hk}/n_h\} = (\bar{T}_{1h}, ..., \bar{T}_{Gh})'$$

Let $\mathbf{R}_h = \mathbf{D}_{\bar{T}_h}^{-1}\bar{\mathbf{y}}_h = [(\bar{y}_{1h}/\bar{T}_{1h}), ..., (\bar{y}_{Gh}/\bar{T}_{Gh})]' = (R_{1h}, ..., R_{Gh})'$ where $\mathbf{D}_{\mathbf{u}}$ is a diagonal matrix with the elements of $\mathbf{u}$ on the diagonal. The $\{R_{gh}\}$ are ratio estimates of incidence densities for numbers of events per unit of time for exposure. Let $\mathbf{a}_{hk} = (\mathbf{y}'_{hk}, T'_{hk})'$ and $\bar{\mathbf{a}}_h = (\bar{\mathbf{y}}'_h, \bar{\mathbf{T}}'_h)'$. An unbiased estimator for the covariance matrix of $\bar{\mathbf{a}}_h$ is $\mathbf{V}_h = \sum_{k=1}^{n_h}(\mathbf{a}_{hk} - \bar{\mathbf{a}}_h)(\mathbf{a}_{hk} - \bar{\mathbf{a}}_h)'/n_h(n_h - 1)$; by Taylor series linearization, a consistent estimator for the covariance matrix of $\mathbf{R}_h$ is $\mathbf{V}_{\mathbf{R}_h} = \mathbf{L}_h\mathbf{V}_h\mathbf{L}'_h$ where $\mathbf{L}_h = \mathbf{D}_{\mathbf{R}_h}[\mathbf{D}_{\bar{\mathbf{y}}_h}^{-1}, -\mathbf{D}_{\bar{\mathbf{T}}_h}^{-1}]$ when sample sizes n_h are sufficiently large (that is, $n_h \geq 60$).

Let $\mathbf{R} = (\mathbf{R}'_1, ..., \mathbf{R}'_H)'$ and let $\mathbf{V}_R$ be the block diagonal matrix with the $\{\mathbf{V}_{\mathbf{R}_h}\}$ as the respective diagonal blocks. Then $\mathbf{V}_R$ is a consistent estimator for the covariance matrix of $\mathbf{R}$. For sufficiently large sample sizes $\{n_h\}$, the vector $\mathbf{R}$ of estimates for incidence densities has an approximately multivariate normal distribution; that is, for each (g, h), at least 10 patients have at least one event and each n_h is at least 30. The vector $\mathbf{R}$ and its estimated covariance matrix $\mathbf{V}_R$ can be computed via SAS/IML (either directly or in combination with PROC MEANS and PROC CORR in Base SAS software.

Loglinear models $\mathbf{R}\hat{=}\{\exp(\mathbf{Xb})\}$, where $\hat{=}$ means is estimated by, can be used to describe the variation among the elements of $\mathbf{R}$ by applying weighted least squares to fit the linear model $\mathbf{Xb}$ to $\mathbf{F} = \mathbf{log}(\mathbf{R})$; for this purpose $\mathbf{X}$ is a $(GH \times t)$ matrix with full rank $t \leq (GH)$, $\mathbf{b}$ is a corresponding vector of estimated parameters and the weights are based on $\mathbf{V}_{\mathbf{F}} = \mathbf{D}_{\mathbf{R}}^{-\mathbf{1}}\mathbf{V}_R\mathbf{D}_R^{-1}$. Thus $\mathbf{b} = (\mathbf{X}'\mathbf{V}_{\mathbf{F}}^{-\mathbf{1}}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}_{\mathbf{F}}^{-\mathbf{1}}\mathbf{F}$ and $\mathbf{V}_{\mathbf{b}} = (\mathbf{X}'\mathbf{V}_{\mathbf{F}}^{-\mathbf{1}}\mathbf{X})^{-1}$ is a consistent estimator for its covariance matrix. When sample sizes are sufficiently large for $\mathbf{R}$ to have an approximately multivariate normal distribution, $\mathbf{b}$ also has an approximately multivariate normal distribution.

A linear hypothesis $\mathbf{Cb}\hat{=}\mathbf{0}$ where $\mathbf{C}$ is a $(c \times t)$ matrix with full rank $c \leq t$ can be tested with $Q_C = \mathbf{b}'\mathbf{C}'(\mathbf{C}\mathbf{V}_b\mathbf{C}')^{-1}\mathbf{Cb}$. This criterion approximately has the chi-square distribution with de-

grees of freedom $(d.f.) = c$ when the hypothesis is true and sample sizes are sufficiently large. Goodness of fit of the model can be evaluated with $Q = (\mathbf{F} - \mathbf{Xb})'\mathbf{V_F^{-1}}(\mathbf{F} - \mathbf{Xb})$. Also, note $Q = \mathbf{F'W'}(\mathbf{WV_FW'})^{-1}\mathbf{WF}$ where $\mathbf{W'}$ is any $GH \times (GH - t)$ orthocomplement to $\mathbf{X}$ (that is, $\mathbf{WX} = \mathbf{0}$). This criterion approximately has the chi-square distribution with $d.f. = (GH - t)$ when the sample sizes are sufficiently large and $\mathbf{F}$ is compatible with the model (that is, $\mathbf{WF} \hat{=} \mathbf{0}$).

Computations for weighted least squares analysis of loglinear models for $\mathbf{R}$ can be made with PROC CATMOD (after input of $\mathbf{F}$ and $\mathbf{V_F}$ from other procedures) or SAS/IML. When the sample sizes $\{n_h\}$ are moderate rather than large, the criterion Q_C can often have more correct evaluation with a F-approximation rather than a chi-square approximation; see Chapter 15.7 of Stokes et al [2000]; that is, apply the F-distribution with $d.f. = (c, (N - c + 1))$ to $\{(N - c + 1)Q_c/N_c\}$ where N expresses $d.f.$ for $\mathbf{V_F}$; that is, $N = \sum_{h=1}^{H}(n_h - 1) = (n - H)$ or $N = (n_0 - 1)$ where n_0 is the sample size for the group which tends to have much larger diagonal elements for $n_h\mathbf{D^{-1}}_{\mathbf{R}_h}\mathbf{V}_{\mathbf{R}_h}\mathbf{D^{-1}}_{\mathbf{R}_h}$ than the other groups. Here, $(N - c + 1)Q_C/Nc$ is a Hotelling T^2 type of transformation for Q_C.

Contact Information:

Maura Stokes, SAS Institute Inc., SAS Campus Drive, Cary, NC 27513 Email: maura.stokes@sas.com