Journal of Biopharmaceutical Statistics, 8(1), 87–102 (1998)

AN APPLICATION OF METHODS FOR CLUSTERED BINARY RESPONSES TO A CARDIOVASCULAR STUDY WITH SMALL SAMPLE SIZE

Gheorghe Luta,[1] *Gary G. Koch,*[1] *Wayne E. Cascio,*[2] *and William T. Smith*[2]

[1]Department of Biostatistics
School of Public Health
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina 27599-7400

[2]Department of Medicine
School of Medicine
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina 27599-7075

Key words. Exact logistic regression; Conditional logistic regression; Multiperiod study; Odds ratios; Confidence intervals

Abstract

This paper discusses statistical methods for a cardiovascular study in which each of eight animals had a dichotomous outcome observed for each of several treatments. There were five treatments in all: shunt, control, two doses of a test drug for potentially causing an unfavorable cardiovascular event, and a combination of the test drug and a counteracting agent. Exact conditional methods were used through LogXact, a statistical software for exact logistic regression and an alternative framework for performing a large class of nonparametric tests performed by StatXact. The results agreed reasonably with asymptotic methods even though the sample size was small.

1. Introduction

The logistic regression model is a statistical model used to describe the relationship between a binary response variable and one or more explanatory variables [see Hosmer and Lemeshow (1)]. Often the sample sizes involved are not large enough to justify the use of the asymptotic statistical methods of inference, so exact conditional methods of inference, as implemented in LogXact (2), are recommended.

Following the terminology used in Mehta and Patel (3,4), we will classify logistic regression models as unstratified (population-averaged model for independent binary responses) or stratified (cluster-specific model for correlated binary responses). An important application of the exact stratified logistic regression model is the situation of clustered binary data consisting of few experimental units, and repeated binary observations for each unit, each experimental unit being managed as a stratum.

The computing environment of LogXact provides results from asymptotic unconditional inference and exact conditional inference for the unstratified logistic regression model. It also provides results from asymptotic conditional inference and exact conditional inference for the stratified logistic regression model. The conditional inference methods (both asymptotic and exact) eliminate the nuisance parameters by conditioning on the observed values of their sufficient statistics. Furthermore, the exact conditional methods utilize the exact permutational distribution of the sufficient statistics for the parameters of interest. It is important to note that a large class of asymptotic and exact nonparametric tests, as performed by StatXact (5), can be expressed in terms of parameters and their sufficient statistics in unstratified and stratified logistic regression models [see, for example, Agresti (6) and Mehta and Patel (4)]. Readers are referred to the Appendix and the original articles for further details.

Other methodologies that account for the correlation structure among the within-cluster observations, namely the generalized estimating equations (GEE) approach for the population-averaged model [see Liang, Zeger, and Qaqish (7)], and the mixed effects approach for the cluster-specific model [see Stiratelli, Laird, and Ware (8)], may not be properly applicable to the small number of primary sampling units and small sample size involved in the experiment which is the focus of this discussion.

The paper shows how LogXact was used for the statistical analysis of a small real data set involving a small number of clusters. Asymptotic and exact methods of statistical inference implemented in LogXact will be illustrated and compared.

2. Application

The main goal of the experimental study to be discussed was the investigation of whether combining a test drug, which potentially causes an unfavorable cardiovas-

cular event, with a counteracting agent reduced the risk of the event to a level similar to that of a control.

The experimental units were eight animals. The experimental method utilized the intracoronary infusion of a drug followed by arrest of coronary flow and the development of regional ischemia. The insertion of the intracoronary artery catheter required arrest of coronary flow and consequently the placement of the intracoronary artery catheter is considered to be the first treatment, and is termed "shunt." Each animal received a sequence of up to four other treatments. The possible treatments consisted of the control (no drugs present), the combination of a low dose (1 unit) of the test drug and the counteracting agent, a low dose (1 unit) of the test drug and a high dose (2 units) of the test drug. Following the administration of each treatment and the arrest of the blood flow in the distribution of the left anterior descending artery, the presence or absence of the adverse cardiovascular event was recorded during an 8-min interval. After the ischemic interval the heart was reperfused for 50 min to establish normal physiological parameters before the next treatment was tested.

The allocation of the treatments to animals was according to sequences which were of interest for shedding light on the extent to which response to a particular treatment was different from the preceding treatment or the subsequent treatment. In this regard, four of the eight animals were given the same sequence: shunt, control, low-dose test drug, the combination of the test drug and the counteracting agent, and control again. This sequence was preferred by experimenters in terms of assessment for whether the response to the second use of control was similar to the response to the first use of control. As a consequence of these considerations, the study is not formally viewed as a crossover study even though its multiperiod structure has a similar nature to what a crossover study might have. Correspondingly, both carryover effects and period effects are assumed to be ignorable. The assumption is supported by the fact that after the ischemic interval each animal was reperfused for 50 min to establish normal physiological parameters before the allocation of the next treatment. In addition, some evaluation for period effects is being provided as part of the analysis.

Since both the low dose and the high dose of the test drug caused the event in the first two animals, the experimenters decided to use only the low dose for the remainder of the study. The control was intended to be given twice to each of the eight animals with the first administration after the shunt and the second administration ending the sequence. Since two animals (the second and the eighth animal) were withdrawn from the experiment before completion, the control treatment was given twice only to the remaining six animals and only once to the ones that were withdrawn. As a consequence, no animal received all five different treatments.

The corresponding data set has a small sample size (37 observations) with a small number of clusters (8 animals) and repeated (up to 5) binary responses in each of them. The outcome of interest was Event, a dichotomous variable, coded as 1 if the cardiovascular event was present, 0 if not. The explanatory variables considered

were: Treatment, with the values: S (shunt), C (control), DA (combination of the test drug and the counteracting agent), D1 (low-dose test drug), and D2 (high-dose test drug); Period, the order of administration of the treatments, with values from P1 to P5; and Animal, with values from A1 to A8. The administered treatment and the dichotomous cardiovascular outcome for each animal and each period are presented in Table 1.

3. Analysis Plan

The small data set that involved a small number of clusters and two "mathematically degenerate" predictors (i.e., the combination of the test drug and the counteracting agent never caused the event, and the high dose of the test drug always did) was analyzed using LogXact, a computing environment that allows one to do exact conditional inference for unstratified and stratified logistic regression models. The analysis strategy involved four planned steps, as follows:

The first step started by considering a stratified logistic regression model with Event as the outcome variable, Animal as the stratification variable, and Treatment coded as an ordinal variable with equally spaced levels (i.e., 0 for S, 1 for C, 2 for DA, 3 for D1, and 4 for D2). Since Period is not included in the model, we are in fact pooling all treatment periods for each animal. Three additional stratified logistic regression models were considered: one in which only treatments other than shunt

Table 1. The Cardiovascular Data Set

Period	1	2	3	4	5
Animal 1	S	C	C	D2	D1
	No	No	No	Yes	Yes
2	S	D2	C	D1	
	No	Yes	No	Yes	
3	S	C	D1	DA	C
	No	Yes	Yes	No	No
4	S	C	D1	DA	C
	No	No	Yes	No	No
5	S	C	DA	D1	C
	Yes	No	No	No	No
6	S	C	D1	DA	C
	No	No	Yes	No	No
7	S	C	D1	DA	C
	No	No	Yes	No	No
8	S	C	D1		
	Yes	Yes	Yes		

were considered, a second one in which both the shunt and the control were excluded, and a third one in which only the shunt, the control, and the combination of the test drug and the counteracting agent were considered. The unstratified counterparts of these four logistic regression models were also applied with LogXact, not as a part of the main analysis, but rather as a way of assessing the effect of stratification for animal.

The second step consisted of the repetition of the first one with the only difference being the coding of Treatment as a classification variable, after pooling together the low and the high dose of the test drug to obtain a group with more adequate sample size from the perspective of statistical power (the high dose was administered only to the first two animals). This step was performed for illustrative purposes and to identify the extent to which less powerful results are obtained.

The third step was based on the fact that the treatments had a structure that identified planned pairwise comparisons and what to expect from them at an $\alpha = .05$ significance level. We expected significant differences for the following pairwise comparisons: C vs. D1 + D2, C + DA vs. D1 + D2, and DA vs. D1 + D2. No significant results were expected for the following pairwise comparisons: S vs. C (verify an assumption), DA vs. S, and DA vs. C. Although these 6 pairwise comparisons were of primary interest, a total of 14 pairwise comparisons were performed for purposes of descriptive completeness. The methods for these comparisons were based on the use of asymptotic score tests, asymptotic conditional score tests, and exact conditional score tests (with the latter being considered the tests of choice) from unstratified and stratified logistic regression models including only the two treatments (or pooling of treatments) to be compared. No adjustment of significance levels were made for these tests because their role was to support supplementary interpretation of the test for Treatment as an ordinal variable. For the main comparisons (i.e., D1 + D2 vs. C, D1 + D2 vs. DA, and D1 + D2 vs. C + DA), estimates and 95% confidence intervals (two-sided) for the corresponding odds ratios were constructed from the unstratified and stratified logistic regression models.

In the last step the effect of the order of administration of treatments (Period) was evaluated by considering a stratified logistic regression model with Event as the outcome variable, Animal as the stratification variable, and Treatment and Period as explanatory variables. Period was coded as having two levels: the pooling of the first three periods (the reference level) and the pooling of the last two periods, by using the indicator variable P45. Since the first period corresponds to S, the second and the fifth periods correspond mainly to C, and the third and the fourth periods correspond to mixtures of C, DA, D1, and D2, the coding is expected to address the extent to which the response during the earlier periods (i.e., 1, 2, 3) might differed from that during the latter periods (i.e., 4, 5); it also avoids any numerical problems (degeneracy due to overconditioning) since C, DA, D1, and D2 occur at least once during periods 2 or 3, and at least once during periods 4 or 5. Treatment was coded as having four levels: S (the reference), C, DA, and the pooling of D1 and

D2, but in addition the codification took into account the dosage of the test drug by assigning a score of 1 for D1, a score of 2 for D2, and a score of 0 for otherwise. The corresponding coding variables were denoted as: C, DA, and D012, respectively. It should be noted that this coding represents an intermediate step between the previous coding of Treatment as an ordinal variable and as a classification variable. For comparative purposes, the unstratified version of this logistic regression model was also applied.

4. Results

For the stratified logistic regression models that take into account the ordinal structure of Treatment, the results of the exact conditional score tests were highly significant for the first two models (including all treatments, and all but shunt), with p-values < 0.002 (see Table 2). For the same models the asymptotic conditional score tests (also fully conditional) provided very similar results although slightly bigger. When only DA, D1, and D2 were considered, a p-value of 0.125 was provided by the exact conditional score test and a borderline value of 0.0455 was given by the asymptotic conditional score test. However, both tests were based on only 4 informative strata (corresponding to the third, fourth, sixth, and seventh animals) with a total of 8 observations, corresponding essentially to a comparison of DA and D1. When comparing S, C, and DA, the results were not significant, supporting similar effects of the shunt, the control, and the combination of the test drug and the counteracting agent with respect to the presence of the cardiovascular event.

The unstratified logistic regression models provided smaller p-values (which reached significance for the exact conditional score test when comparing DA, D1, and D2) for the corresponding asymptotic score test and exact conditional score test. This pattern suggests that the lost information due to noninformative animals reduces power to an extent that overcomes any potential gain in power due to stratification for animal. It should also be mentioned that the results of the asymptotic score test

Table 2. Results of Tests for Association of Treatment (Ordinal Variable) and Event[a]

	Unstratified model		Stratified model	
Treatment	ST	ECST	CST	ECST
S C DA D1 D2	.001150	.001190	.001705	.001691
C DA D1 D2	.000233	.000189	.001232	.000868
DA D1 D2	.002169	.001598	.045500	.125000
S C DA	.218571	.274473	.317311	.625000

[a] p-values from the asymptotic score test (ST), asymptotic conditional score test (CST), and exact conditional score test (ECST).

and the exact conditional score test were relatively similar to one another in most cases for the unstratified analysis.

When Treatment was considered as a classification variable, the stratified logistic regression models provided similar results as when considered as an ordinal variable, but with smaller p-values for the first two models and higher p-values for the last model (see Table 3). The unstratified counterparts supported the same conclusions. The asymptotic score tests and the exact conditional score tests from these models provided smaller p-values than their stratified versions, due to the effect of the tradeoff between losing information by ignoring the noninformative animals and taking advantage of directly observable within-cluster effects.

For the planned comparisons of S vs. C, S vs. DA, and C vs. DA, the p-values provided by the exact conditional score tests from the stratified logistic regression models were 1.000 (see Table 4), which even after considering the potential conservativeness of these tests, provides reasonable support for the pairwise similarity of S, C, and DA. When comparing D1 + D2 with C, and D1 + D2 with C + DA, the results of the exact tests were significant at $\alpha = .05$, with corresponding p-values of 0.0014 and 0.0004, respectively. However the pairwise comparison between DA and the pooling of D1 and D2, which is in fact a pairwise comparison between DA and D1, was nonsignificant, p-value $= 0.125$, as noted before.

From a different perspective, the results of assessing each of the 14 pairwise comparisons with 4 types of score tests provided the same conclusion at a .05 significance level for 10 of them. In the remaining 4 situations, the disagreement was partially due to the fact that the exact conditional score test for the stratified model was nonsignificant. There was also a general tendency for the p-values to increase in the following order: the asymptotic score test, the exact conditional score test from the unstratified model, the asymptotic conditional score test, and finally the exact conditional score test for the stratified model.

It can be seen from Table 5 that for the three pairwise comparisons, the upper bound of the 95% exact confidence intervals was infinity when exact estimation was done by using stratified logistic regression models. For the association between treatment (D1 + D2 vs. C + DA) and the presence of the unfavorable cardiovascular

Table 3. Results of Tests for Association of Treatment (Classification Variable) and Event[a]

	Unstratified model		Stratified model	
Treatment	ST	ECST	CST	ECST
S C DA D1 + D2	.000280	.000103	.000989	.001035
C DA D1 + D2	.000130	.000046	.000730	.000434
DA D1 + D2	.000796	.001998	.045500	.125000
S C DA	.465269	.514986	.606531	1.00000

[a] p-values from the asymptotic score test (ST), asymptotic conditional score test (CST), and exact conditional score test (ECST).

Table 4. Results of Pairwise Tests for Association Between Treatment and Event[a]

Treatment		Unstratified model		Stratified model	
		ST	ECST	CST	ECST
S	C	.5308	.6019	.6171	1.000
	DA	.2242	.4872	.3173	1.000
	C+DA	.3337	.5583	.4142	.4375
	D1	.0117	.0406	.0588	.1250
	D2	.0528	.1333	.1573	.5000
	D1+D2	.0049	.0128	.0295	.0486
C	DA	.3716	.5906	.4795	1.000
	D1	.0008	.0015	.0027	.0041
	D2	.0088	.0500	.0896	.1667
	D1+D2	.0002	.0005	.0009	.0014
DA	D1	.0021	.0047	.0455	.1250
	D2	.0082	.0476	—	—
	D1+D2	.0008	.0020	.0455	.1250
C+DA	D1+D2	.0001	.0001	.0002	.0004

[a] *p*-values from the asymptotic score test (ST), asymptotic conditional score test (CST), and exact conditional score test (ECST).

Table 5. The Association Between Selected Pairs of Treatments and Event[a]

Estimation	D1+D2 vs. C	D1+D2 vs. DA	D1+D2 vs. C+DA
Unstratified			
Asymptotic	54.0 (4.2, 692.5)	Undefined	76.5 (6.0, 963.1)
Exact[b]	40.6 (3.2, 2552.8)	27.2 (2.6, ∞)	57.2 (4.7, 3535.4)
Stratified			
Asymptotic	Undefined	Undefined	Undefined
Exact[b]	16.9 (2.3, ∞)	5.3 (0.6, ∞)	21.7 (3.0, ∞)

[a] Odds ratio estimate (OR) and 95% confidence interval (95% CI) for the OR.

[b] Median unbiased estimate is used when the asymptotic estimate is undefined.

event the 95% exact confidence intervals indicate that the odds ratio is at least 3 for the cluster-specific effect, and at least 4.7 for the population-averaged effect. When not undefined, the estimates become smaller (attenuated) and the corresponding confidence intervals become wider when moving from the unstratified asymptotic results to the stratified exact results (however, the unstratified asymptotic results and the unstratified exact results were similar).

Both the stratified and the unstratified exact logistic regression models presented in Table 6 supported the hypothesis of no difference with respect to the odds of presence of the event between C and S, between DA and S, and also between the last two and the first three periods of administration of treatments. It is important to note that for the comparison of DA and S the median unbiased estimates of the odds

Table 6. Results from Exact Logistic Regression Models[a]

	Unstratified model		Stratified model	
Variable	OR (95% CI)	p-value	OR (95% CI)	p-value
D012	16.2 (1.9, ∞)	.0040	8.0 (1.2, ∞)	.0326
DA	3.5 (0, 136.4)	1.000	7.0 (0, 272.8)	1.000
C	.9 (.04, 15.5)	1.000	.9 (.01, 68.7)	1.000
P45	.3 (0, 2.9)	.2170	.5 (0, 5.6)	.5600

[a]Odds ratio estimate (OR), 95% confidence interval (95% CI) for the odds ratio, and p-value from exact conditional score test for testing odds ratio equal to 1.

ratios were 3.5 (unstratified model) and 7 (stratified model) while the lower bounds of the corresponding 95% exact confidence intervals were 0. In a situation like this it is recommended to rely more on confidence intervals and less on point estimates when making inferences [see Mehta and Patel (4)]. In both models there was a significant effect per 1 unit increase in dose of the test drug when compared with S. It should be mentioned that the latter effect was estimated as being greater and more significant in the unstratified model (however, in both models the upper bound of the 95% exact confidence interval was infinity). The results were consistent with previous interpretations of the effects of stratification for animal. It is important to note that the effects of stratification for this small data set (37 observations) involving a small number of cluster (8) were similar to those found by Ten Have, Landis, and Weaver (9) when comparing methods for clustered binary data on a larger data set (131 observations) involving a moderate number of clusters (31).

It should be mentioned that for the models presented in Table 6 only the exact conditional inference could be done, and that even this method failed (i.e., degenerate permutational distributions) when larger models were tried. Both the asymptotic conditional method and the exact conditional method were possible for a stratified model including C and D012. The contributions of DA and P45 were assessed to evaluate their addition to the model, one variable at a time. The asymptotic conditional score tests and the exact conditional score tests provided very similar results for the variables in the model, and they also supported the conclusion of no need for adding DA and P45 to the model. The corresponding odds ratio estimates and 95% confidence intervals were similar (see Table 7).

5. Conclusion

The results of the statistical analysis support the hypothesis that the combination of the test drug and the counteracting agent may reduce the risk of the unfavorable cardiovascular event to a level comparable to that of a control or shunt. Exact conditional methods were used through the LogXact computing procedures, the

Table 7. Results from Stratified Logistic Regression Models

	Conditional asymptotic inference		Conditional exact inference	
Variable	OR (95% CI)	p-value	OR (95% CI)	p-value
D012[a]	19.2 (1.5, 240.6)	.0033	15.9 (1.7, 904.1)	.0029
C[a]	.97 (.05, 16.1)	.9802	.96 (.01, 78.2)	1.000
DA[b]		.3628		1.000
P45[b]		.1132		.3261

[a] Odds ratio estimate (OR), 95% confidence interval (95% CI) for the odds ratio, and p-value from conditional score test for testing odds ratio equal to 1, from a model that includes D012 and C.

[b] p-values are from conditional score tests for entry of variable to the model that includes D012 and C.

software of choice for small data sets involving a small number of clusters (or strata), and also an alternative framework for performing a large class of asymptotic and exact nonparametric tests performed by StatXact. It should be noticed that the results of the exact conditional methods agreed reasonably with the corresponding results of the asymptotic methods even though the sample size was small. The advantage of the exact results is that their interpretation can proceed without the awkward caution that often accompanies asymptotic results.

Acknowledgments

The work of Wayne E. Cascio was supported by P01 HL27430, RO1 HL48769 from the NIH, Heart, Lung, and Blood Institute, Bethesda, Maryland. The work of William T. Smith was supported by a Howard Hughes Medical Institute Medical Student Research Training Fellowship Grant.

References

1. Hosmer DW, Lemeshow S: *Applied Logistic Regression.* Wiley, New York, 1989.
2. Cytel Software Corporation: *LogXact. Software for Exact Logistic Regression.* Cytel Software Corporation, Cambridge, MA, 1992.
3. Mehta CR, Patel NR: Exact logistic regression: Theory and examples. *Statist Med* 14:2143–2160, 1995.
4. Mehta CR, Patel NR: *LogXact-Turbo User Manual.* Cytel Software Corporation, Cambridge, MA, 1993.
5. Cytel Software Corporation: *StatXact. Statistical Software for Exact Nonparametric Inference.* Cytel Software Corporation, Cambridge, MA, 1989.
6. Agresti A: A survey of exact inference for contingency tables. *Statist Sci* 7(1):131–177, 1992.
7. Liang K-Y, Zeger SL, Qaqish B: Multivariate regression analysis for categorical data. *J Roy Statist Soc Ser B* 54:3–40, 1992.

8. Stiratelli R, Laird NM, Ware JH: Random-effects models for serial observations with binary response. *Biometrics* 40:961–971, 1984.
9. Ten Have TR, Landis JR, Weaver SL: Association models for periodontal disease progression: A comparison of methods for clustered binary data. *Statist Med* 14:413–429, 1995.
10. McCullagh P, Nelder JA: *Generalized Linear Models*, 2nd ed. Chapman and Hall, London, 1989.
11. Wedderburn RWM: On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. *Biometrika* 63:27–32, 1976.
12. Nelder JA, Wedderburn RWM: Generalized linear models. *J Roy Statist Soc Ser A* 135:370–384, 1972.
13. Sen PK, Singer JM: *Large Sample Methods in Statistics: An Introduction with Applications.* Chapman and Hall, London, 1993.
14. Cox DR, Hinkley DV: *Theoretical Statistics.* Chapman and Hall, London, 1974.

Appendix

The purpose of this Appendix is to summarize aspects of the statistical theory for the methods that were applied to the cardiovascular study in this paper. It is provided so that the reader has it conveniently available for helping their understanding of the applications that the paper provides. The discussion is substantially based on Mehta and Patel (3,4).

A.1 The Logistic Regression Model

Consider a set of independent binary random variables $Y_1, Y_2, \ldots, Y_n$ and let $y_1, y_2, \ldots, y_n$ be their observed values. Let $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{it})'$ be the $t \times 1$ vector of explanatory variables and $\pi(\mathbf{x}_i) = P(Y_i = 1 \mid \mathbf{x}_i)$, $i = 1, 2, \ldots, n$. The logistic regression model can be represented as:

$$\log\left(\frac{\pi(\mathbf{x}_i)}{1-\pi(\mathbf{x}_i)}\right) = \mathbf{x}_i'\boldsymbol{\beta} \tag{A.1}$$

where $\boldsymbol{\beta}$ represents the t × 1 vector of unknown parameters and $i = 1, 2, \ldots, n$. If we denote by $\boldsymbol{\pi}$ the n × 1 vector with elements $\pi(\mathbf{x}_i)$ and by $\mathbf{X}$ the n × t matrix having $\mathbf{x}_i'$ as rows, (A.1) can be expressed as:

$$\text{logit}(\boldsymbol{\pi}) = \mathbf{X}\boldsymbol{\beta} \tag{A.2}$$

We will partition the t × 1 vector $\boldsymbol{\beta}$ into three components: $\boldsymbol{\beta}_0$, the s × 1 vector of stratum specific intercepts (indicator variables) corresponding to the s strata, with s = 1 in the unstratified case; $\boldsymbol{\beta}_1$, a (t-s-d) × 1 vector and $\boldsymbol{\beta}_2$ a d × 1 vector of parameters that describe variation across strata. Without loss of generality we will manage $\boldsymbol{\beta}_2$ as the vector of parameters of interest and $\boldsymbol{\beta}_1$ as a vector of nuisance parameters. Accordingly we will partition $\mathbf{X}$ into $\mathbf{X}_0$, $\mathbf{X}_1$, and $\mathbf{X}_2$ to match the partition of $\boldsymbol{\beta}$. In this way, we have a unified way of addressing both the unstratified and the stratified logistic regression models.

A.2 Unconditional Asymptotic Inference

This method of inference is based on maximizing the unconditional likelihood:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \{\pi(\mathbf{x}_i)^{y_i}[1 - \pi(\mathbf{x}_i)]^{1-y_i}\} \tag{A.3}$$

or equivalently by maximizing the log likelihood $l(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta})$. The estimating function and the Fisher information matrix are as follows:

$$\mathbf{U}(\boldsymbol{\beta}) = \left(\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_j}\right) = \left(\sum_{i=1}^{n} x_{ij}[y_i - \pi(\mathbf{x}_i)]\right) \tag{A.4}$$

$$\mathbf{I}(\boldsymbol{\beta}) = E\left(-\left(\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_k}\right)\right) = \left(\sum_{i=1}^{n} x_{ij}x_{ik}\pi(\mathbf{x}_i)[1 - \pi(\mathbf{x}_i)]\right) \tag{A.5}$$

It should be noted that the elements of the matrix of the second-order partial derivatives of the log likelihood do not depend on the observations $y_1, \ldots, y_n$. In fact this happens for all generalized linear models that use the canonical link [see McCullagh and Nelder (10)]. The ML estimating equations are:

$$\sum_{i=1}^{n} x_{ij}[y_i - \hat{\pi}(\mathbf{x}_i)] = 0, \quad j = 1, \ldots, t \tag{A.6}$$

Under general conditions (11), the ML estimates of the model parameters exist and are unique. Solving the estimating equations is usually accomplished with the Newton-Raphson algorithm, a simplification of the Fisher scoring algorithm for the generalized linear models using the canonical link [see Nelder and Wedderburn (12)].

Asymptotic methods such as the likelihood ratio statistic, the Wald statistic, and the score statistic can be used to perform hypothesis tests of H_0:$\boldsymbol{\beta}_2 = \mathbf{0}$ vs. H_a:$\boldsymbol{\beta}_2 \neq \mathbf{0}$. Let $\hat{\boldsymbol{\beta}}$ be the MLE of $\boldsymbol{\beta}$ from model (A.2) and let $\tilde{\boldsymbol{\beta}}$ be the MLE under the restriction that $\boldsymbol{\beta}_2 = \mathbf{0}$. The likelihood ratio statistic is:

$$Q_l = 2\{l(\hat{\boldsymbol{\beta}}) - l(\tilde{\boldsymbol{\beta}})\} \tag{A.7}$$

The Wald statistic is:

$$Q_w = \hat{\boldsymbol{\beta}}_2'[\mathbf{V}_{22}(\hat{\boldsymbol{\beta}})]^{-1}\hat{\boldsymbol{\beta}}_2 \tag{A.8}$$

where $\mathbf{V}_{22}(\hat{\boldsymbol{\beta}})$ is the lower right corner of $\mathbf{V}(\hat{\boldsymbol{\beta}}) = [\mathbf{I}(\hat{\boldsymbol{\beta}}]^{-1}$. The score statistic is:

$$Q_s = \mathbf{U}'(\tilde{\boldsymbol{\beta}})[\mathbf{I}(\tilde{\boldsymbol{\beta}})]^{-1}\mathbf{U}(\tilde{\boldsymbol{\beta}}) \tag{A.9}$$

Compared with the previous two tests, the score test is computationally easier and it does not require determination of $\hat{\boldsymbol{\beta}}$. With large enough sample size for the sufficient statistics for $\boldsymbol{\beta}$ to have an approximately multivariate normal distribution, all three of these statistics have an asymptotic chi-squared distribution with d degrees of freedom under H_0 [see Sen and Singer (13)].

A $100(1-\alpha)\%$ asymptotic confidence interval for univariate β_2 is constructed by "inverting" the hypothesis test based on the Wald statistic:

$$\hat{\beta}_2 \pm z_{1-\alpha/2}\left[V_{22}(\hat{\boldsymbol{\beta}})\right]^{1/2} \tag{A.10}$$

where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)$ percentile of the standard normal distribution.

A.3 Conditional Asymptotic Inference

This method of inference is based on maximizing the conditional likelihood obtained by conditioning (A.3) on the observed values of the sufficient statistics for $\boldsymbol{\beta}_0$, the vector of intercepts for the s strata. Following McCullagh and Nelder (10), the observed value of the sufficient statistic T_j for β_j is:

$$t_j = \sum_{i=1}^{n} x_{ij} y_i, \quad j = 1, \ldots, t \tag{A.11}$$

Denote by $\mathbf{T}_0$, $\mathbf{T}_1$, and $\mathbf{T}_2$ the vectors of sufficient statistics corresponding to $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, and $\boldsymbol{\beta}_2$. It follows that the conditional probability density function of $\mathbf{T}_1$ and $\mathbf{T}_2$ given $\mathbf{T}_0 = \mathbf{t}_0$, the vector of observed number of responses in each stratum, is:

$$f_{\beta_1,\beta_2}(\mathbf{t}_1, \mathbf{t}_2 \mid \mathbf{t}_0) = \frac{C(\mathbf{t}_0, \mathbf{t}_1, \mathbf{t}_2)\exp(\mathbf{t}_1'\boldsymbol{\beta}_1 + \mathbf{t}_2'\boldsymbol{\beta}_2)}{\sum_{\mathbf{u}_1,\mathbf{u}_2} C(\mathbf{t}_0, \mathbf{u}_1, \mathbf{u}_2)\exp(\mathbf{u}_1'\boldsymbol{\beta}_1 + \mathbf{u}_2'\boldsymbol{\beta}_2)} \tag{A.12}$$

where $C(\mathbf{t}_0, \mathbf{u}_1, \mathbf{u}_2)$ is the count of the number of elements of the set:

$$\{\mathbf{y} : \mathbf{X}_0'\mathbf{y} = \mathbf{t}_0, \mathbf{X}_1'\mathbf{y} = \mathbf{u}_1, \mathbf{X}_2'\mathbf{y} = \mathbf{u}_2\},$$

a count of binary sequences of length n subject to appropriate restrictions, which imply that the summation in the denominator involves only terms with $C(\mathbf{t}_0, \mathbf{u}_1, \mathbf{u}_2) \geq 1$.

The conditional probability density function from (A.12) is the conditional likelihood function of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ given $\mathbf{T}_0 = \mathbf{t}_0$. The elimination of $\boldsymbol{\beta}_0$ from the likelihood function by conditioning on the observed value of its sufficient statistic can be justified by the fact that $\boldsymbol{\beta}_2$ is the vector of parameters of interest. Furthermore, if the dimension of $\boldsymbol{\beta}_0$ is large relative to the number of observations, the MLEs

for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ may be inconsistent whereas the conditional MLEs can be consistent [see Cox and Hinkley (14)].

In addition to conditional MLEs, asymptotic conditional tests (conditional likelihood ratio statistic, conditional Wald statistic, and conditional score statistic) for H_0:$\boldsymbol{\beta}_2 = \mathbf{0}$ vs. $H_a : \boldsymbol{\beta}_2 \neq \mathbf{0}$, and also asymptotic confidence intervals for scalar parameters, can be obtained similarly to the case of unconditional likelihood inference. For example, the conditional score statistic is defined as:

$$\mathbf{Q}_{cs} = \mathbf{U}'(\tilde{\boldsymbol{\beta}} \mid \mathbf{t}_0)\left[\mathbf{I}(\tilde{\boldsymbol{\beta}} \mid \mathbf{t}_0)\right]^{-1}\mathbf{U}(\tilde{\boldsymbol{\beta}} \mid \mathbf{t}_0), \tag{A.13}$$

being based on the conditional likelihood from (A.12). It is assumed that the sample size is large enough to support an approximate multivariate normal distribution for the sufficient statistics $\mathbf{T}_1$ and $\mathbf{T}_2$, so asymptotically the three conditional tests are chi-squared distributed with d degrees of freedom, under H_0.

A.4 Conditional Exact Inference

This method of inference is based on the generation of the conditional permutational distribution for the sufficient statistic of the parameter(s) of interest. Its use is recommended when one has small samples or unbalanced data sets or highly stratified data, since in these cases the asymptotic methods discussed previously may not be valid. Similarly to (A.12), the conditional probability density function of $\mathbf{T}_2$ given $\mathbf{T}_1 = \mathbf{t}_1$ and $\mathbf{T}_0 = \mathbf{t}_0$ is:

$$f_{\boldsymbol{\beta}_2}(\mathbf{t}_2 \mid \mathbf{t}_1, \mathbf{t}_0) = \frac{C(\mathbf{t}_0, \mathbf{t}_1, \mathbf{t}_2)\exp(\mathbf{t}_2'\boldsymbol{\beta}_2)}{\sum_{\mathbf{u}_2} C(\mathbf{t}_0, \mathbf{t}_1, \mathbf{u}_2)\exp(\mathbf{u}_2'\boldsymbol{\beta}_2)} \tag{A.14}$$

where $C(\mathbf{t}_0, \mathbf{t}_1, \mathbf{u}_2)$ is the count of the number of elements of the set:

$$\{\mathbf{y} : \mathbf{X}_0'\mathbf{y} = \mathbf{t}_0, \mathbf{X}_1'\mathbf{y} = \mathbf{t}_1, \mathbf{X}_2'\mathbf{y} = \mathbf{u}_2\},$$

and the summation in the denominator involves only terms with $C(\mathbf{t}_0, \mathbf{t}_1, \mathbf{u}_2) \geq 1$.

To test H_0:$\boldsymbol{\beta}_2 = \mathbf{0}$ vs. H_a:$\boldsymbol{\beta}_2 \neq \mathbf{0}$, we can use the exact conditional probability density function of $\mathbf{T}_2$ under H_0, namely:

$$f_0(\mathbf{t}_2 \mid \mathbf{t}_1, \mathbf{t}_0) = \frac{C(\mathbf{t}_0, \mathbf{t}_1, \mathbf{t}_2)}{\sum_{\mathbf{u}_2} C(\mathbf{t}_0, \mathbf{t}_1, \mathbf{u}_2)} \tag{A.15}$$

One of the useful exact conditional test statistics is the exact conditional score statistic. Let $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_2(\mathbf{t}_1, \mathbf{t}_0)$ and $\mathbf{V}_2 = \mathbf{V}_2(\mathbf{t}_1, \mathbf{t}_0)$ be the respective conditional mean vector and the conditional variance-covariance matrix of $\mathbf{T}_2$ given $\mathbf{T}_1 = \mathbf{t}_1$

and $\mathbf{T}_0 = \mathbf{t}_0$, under H_0:$\boldsymbol{\beta}_2 = \mathbf{0}$. We can define the exact conditional score statistic as:

$$Q_{ecs} = (\mathbf{T}_2 - \boldsymbol{\mu}_2)'\mathbf{V}_2^{-1}(\mathbf{T}_2 - \boldsymbol{\mu}_2) \tag{A.16}$$

and its observed value:

$$q_{ecs} = (\mathbf{t}_2 - \boldsymbol{\mu}_2)'\mathbf{V}_2^{-1}(\mathbf{t}_2 - \boldsymbol{\mu}_2) \tag{A.17}$$

The corresponding p-value is:

$$p = \sum_{\mathbf{u}\in\mathcal{R}} f_0(\mathbf{u} \mid \mathbf{t}_0, \mathbf{t}_1) \tag{A.18}$$

with the rejection region $\mathcal{R} = \{\mathbf{u} : (\mathbf{u} - \boldsymbol{\mu}_2)'\mathbf{V}_2^{-1}(\mathbf{u} - \boldsymbol{\mu}_2) \geq q_{ecs}\}$.

The exact conditional score statistic is an exact analog of the asymptotic conditional score test. However, it should be noted that the former test is fully conditional and uses an exact variance-covariance matrix as opposed to the latter one which is only (in general) partially conditional and uses an asymptotic variance-covariance matrix.

Exact parameter estimation is performed only for scalar parameters. If the (fully) conditional likelihood cannot be maximized (i.e, the observed value of the sufficient statistic for the parameter of interest is at one extreme of its range: $\mathbf{t}_2 = \mathbf{t}_{2,\min}$ or $\mathbf{t}_2 = \mathbf{t}_{2,\max}$), a median unbiased estimate is used instead. This point estimate, denoted by $\bar{\beta}_2$, satisfies:

$$f_{\bar{\beta}_2}(t_2 \mid \mathbf{t}_1, \mathbf{t}_0) = 0.50 \tag{A.19}$$

A $100(1-\alpha)\%$ two-sided exact confidence interval for β_2 can also be constructed in an "inverting hypothesis test manner". For $\mathbf{t}_2 \neq \mathbf{t}_{2,\min}$ its lower bound β_{2L} satisfies the condition:

$$\sum_{u\geq t_2} f_{\beta_{2L}}(u \mid \mathbf{t}_0, \mathbf{t}_1) = \alpha/2 \tag{A.20}$$

If $t_2 = t_{2,\min}$, β_{2L} is taken to be $-\infty$. For $t_2 \neq t_{2,\max}$ its upper bound β_{2U} satisfies:

$$\sum_{u\leq t_2} f_{\beta_{2U}}(u \mid \mathbf{t}_0, \mathbf{t}_1) = \alpha/2 \tag{A.21}$$

If $t_2 = t_{2,\max}$, β_{2U} is taken to be ∞. This method produces an interval that has at least the desired coverage [see Mehta and Patel (3)].

A.5 Inference on Contingency Tables Using Logistic Regression Models

Following Agresti (6), a large class of exact nonparametric tests for contingency tables involving a dichotomous response variable can be expressed in terms of parameters and their sufficient statistics in unstratified and stratified logistic regression models. Simple cases, relevant to our study, are presented to illustrate the use of LogXact to perform a broad spectrum of asymptotic and exact nonparametric tests as done by StatXact. For additional examples see also Mehta and Patel (4).

In unstratified logistic regression models involving one categorical explanatory variable, the exact conditional score test for its contribution to the model is equivalent to the exact Pearson chi-squared test (i.e., Fisher's exact test based on the Pearson chi-squared test statistic), to the exact chi-squared test for general association (an alternative to the Fisher-Freeman-Halton test), or to the exact Cochran-Armitage trend test, depending on the explanatory variable being a dichotomous variable, a classification variable, or an ordinal variable with equally spaced levels. The corresponding asymptotic score test is equivalent to the Pearson chi-squared test, to the chi-squared test for general association, or to the Cochran-Armitage trend test.

In stratified logistic regression models involving one categorical explanatory variable, the exact conditional score test for its contribution to the model is equivalent to a stratified version of the exact Pearson chi-squared test for independence, to a stratified version of the exact chi-squared test for general association, or to a stratified version of the exact Cochran-Armitage trend test, depending on the explanatory variable being a dichotomous variable, a classification variable, or an ordinal variable with equally spaced levels. The corresponding asymptotic conditional score test is equivalent to the Mantel-Haenszel chi-squared test, the extended Mantel-Haenszel chi-squared test for general association, or to a stratified version of the Cochran-Armitage trend test.