BIOMETRICS 58, 398–402
June 2002

Testing Trend for Count Data with Extra-Poisson Variability

Erni Tri Astuti[1,2] and Takashi Yanagawa[2,*]

[1]Institute of Statistics, Jakarta Timur 13330, Indonesia
[2]Graduate School of Mathematics, Kyushu University, Fukuoka 812-8581, Japan
*email: yanagawa@math.kyushu-u.ac.jp

SUMMARY. Trend tests for monotone trend or umbrella trend (monotone upward changing to monotone downward or vise versa) in count data are proposed when the data exhibit extra-Poisson variability. The proposed tests, which are called the GS1 test and the GS2 test, are constructed by applying an orthonormal score vector to a generalized score test under an rth-order log-linear model. These tests are compared by simulation with the Cochran–Armitage test and the quasi-likelihood test of Piegorsch and Bailer (1997, *Statistics for Environmental Biology and Toxicology*). It is shown that the Cochran–Armitage test should not be used under the existence of extra-Poisson variability; that, for detecting monotone trend, the GS1 test is superior to the others; and that the GS2 test has high power to detect an umbrella response.

KEY WORDS: Cochran–Armitage test; Generalized score test; Negative binomial distribution; Orthonormal score vector; Toxicology.

1. Introduction

Environmental data often exhibit extra-Poisson variability. For example, Table 1 summarizes data from a study of toxic reproductive response in the aquatic organism *Ceriodaphnia dubia* to the herbicide nitrofen reported in Bailer and Oris (1993). Offspring counts from exposed females are used as a measure of reproductive stress. The table shows that the estimated variances are substantially larger than the estimated means at the two highest exposure levels, an indication of extra-Poisson variability. A classical approach to the problem is to treat the Poisson means as latent variables that are sampled from a gamma distribution (Margolin, Kaplan, and Zeiger, 1981). There are various other approaches, such as using the mean and variance structure implied by the mixed Poisson model (Williams, 1982; Breslow, 1984, 1990). Breslow (1990) developed two versions of the Wald and score tests—one calculated from the assumed mean and variance structure and other using an empirical covariance matrix. Boos (1992) generalized the latter by introducing the generalized score test. Piegorsch and Bailer (1997) developed a test using the empirical variance, which is called the QL test in this article because it was motivated from a quasi-likelihood perspective. Compared with the test that uses only the mean and variance structure, the test that uses the empirical variance has closed form, is particularly simple to compute, and remains valid even when the mean and variance structure is incorrect.

In addition, environmental response data sometimes exhibit an umbrella trend, such as an upward trend with a downturn at high doses or vise versa. For example, Table 2 displays mutagenic response data from a *Salmonella* assay of the chemical Acid Red 114 given in Simpson and Margolin (1986). Six dose levels are used and each dose is replicated three times. The mean responses show that the dose–response in each replicate increases over low doses but then has a downturn in higher doses. Simpson and Margolin (1986) developed a Jonckheere–Terpstra-type recursive test sensitive to that alternative. This issue is also discussed under umbrella alternatives (Neuhäuser et al., 2000).

In this article, we develop trend tests for count data with extra-Poisson variability for data such as given in Tables 1 and 2. The proposed tests are constructed by applying an orthonormal dose vector to the generalized score test of Boos (1992) under an rth order log-linear model. When $r = 1$, the test, which is called the GS1 test, is similar to the QL test, but it is shown that the GS1 test is more faithful to the nominal test level than the QL test. When $r = 2$, it is indicated by sim-

Table 1
Offspring counts for C. dubia *exposed to nitrofen*[a]

Dose (μg/liter)	Number of offspring					Mean	Variance
Control	27	32	34	33	36		
	34	33	30	24	31	31.4	12.93
80	33	33	35	33	36		
	26	27	31	32	29	31.5	10.72
160	29	29	23	27	30		
	31	30	26	29	29	28.3	5.57
235	23	21	7	12	27		
	16	13	15	21	17	17.2	34.84
310	6	6	7	0	15		
	5	6	4	6	5	6.0	13.78

[a] Data from Piegorsch and Bailer (1997, p. 220).

Table 2

Numbers of revertant colonies for Acid Red 114, TA98[a]

	Replicate 1			Replicate 2			Replicate 3		
Dose (ppm)	1	2	3	4	5	6	7	8	9
0	22	23	35	19	17	16	23	22	14
100	60	59	54	15	25	24	27	23	21
333	98	78	50	26	17	31	28	37	35
1000	60	82	59	39	44	30	41	37	43
3333	22	44	33	33	26	23	28	21	30
10,000	23	21	25	10	8	—	16	19	13

[a] Data from Simpson and Margolin (1986).

ulation that the proposed test has higher power in detecting an umbrella response than the GS1 test.

2. Trend Tests

Consider a dose–response experiment with dose levels $d_1, d_2, \ldots, d_k$ $(d_1 < d_2 < \cdots < d_k)$. Suppose that m_i independent counts Y_{ij} are observed, with mean rate of response μ_i at d_i, $i = 1, 2, \ldots, k$. The object of this article is to test the hypothesis $\mu_1 = \mu_2 = \cdots = \mu_k$ against a monotone trend or umbrella trend in μ's. We formulate this problem by introducing a score vector $\mathbf{a_s} = (a_{s1}, a_{s2}, \ldots, a_{sk})'$, to be defined below, and by representing the μ_i's as

$$\log \mu_i = \sum_{s=0}^{r} \beta_s a_{si}, \tag{1}$$

where r is an integer $(r < k)$ and $a_{01} = a_{02} = \cdots = a_{0k}$. Inclusion of arbitrary covariates in this formula is straightforward, but we consider only the factor directly related to dose since our goal is to derive tests for trend.

Now we introduce orthonormal score vectors. Let a dot in a subscript denote the summation over that subscript, e.g., $m_\cdot = \Sigma_i m_i, Y_{i\cdot} = \Sigma_j Y_{ij}$. Define $\mathbf{c_1} = (c_1, c_2 \ldots, c_k)'$, where $c_i = d_i - \bar{d}$ and $\bar{d} = \Sigma_i d_i m_i / m_\cdot$, so that $\Sigma_i c_i m_i = 0$. Also define $\mathbf{c_s} = (c_{s1}, c_{s2}, \ldots, c_{sk})'$, where $c_{si} = c_i^s$ (sth power of c_i) for $s = 1, 2, \ldots, r$ and $\mathbf{c_0} = (1, 1, \ldots, 1)'$. Define the inner product of two vectors as $(\mathbf{a}, \mathbf{b}) = \Sigma_i a_i b_i m_i$ and $\|\mathbf{a}\|^2 = (\mathbf{a}, \mathbf{a})$. Let $\mathbf{a_0}, \mathbf{a_1}, \ldots, \mathbf{a_r}$ be orthonormal vectors obtained by applying the Gram–Schimdt orthonormalization to these vectors, i.e., $\mathbf{a_0} = \mathbf{c_0}/\|\mathbf{c_0}\|$, $\mathbf{d_s^*} = \mathbf{c_s} - \Sigma_{h=0}^{s-1} (\mathbf{c_s}, \mathbf{a_h})\mathbf{a_h}$, $\mathbf{a_s} = \mathbf{d_s^*}/\|\mathbf{d_s^*}\|$. Then $(\mathbf{a_s}, \mathbf{a_l}) = 1$ if $s = l$ and is zero otherwise and $\|\mathbf{a_s}\| = 1$ for all $s = 0, 1, \ldots, r$. We call $\mathbf{a_s}$ the orthonormal score vector.

Let $\boldsymbol{\beta}_{(2)} = (\beta_1, \ldots, \beta_r)'$. Using the orthonormal score vectors, it is shown in the Appendix that the generalized score test (Boos, 1992) for testing H_0: $\boldsymbol{\beta}_{(2)} = 0$ against H_1: $\boldsymbol{\beta}_{(2)} \neq \mathbf{0}$ is given by $GSr = \tilde{S}'_{(2)} \tilde{D}^{-1}_{Y(22)} \tilde{S}_{(2)}$, where $\tilde{S}_{(2)} = (\Sigma_i a_{1i} Y_{i\cdot}, \ldots, \Sigma_i a_{ri} Y_{i\cdot})'$, $\tilde{D}_{Y(22)} = (\Sigma_i \Sigma_j (Y_{ij} - \bar{Y})^2 a_{ti} a_{ui})_{r\times r}$, $Y_{i\cdot} = \Sigma_{j=1}^{m_i} Y_{ij}$, $\bar{Y} = \Sigma_{i=1}^{k} Y_{i\cdot}/m_\cdot$, and $m_\cdot = \Sigma_{i=1}^{k} m_i$. It may be shown that GSr follows a chi-square distribution with r d.f. under H_0 asymptotically when $m_i \to \infty$, $i = 1, 2, \ldots, k$.

Putting $\bar{d} = \Sigma_i^k d_i m_i / m_\cdot$, when $r = 1$, the GSr is written as

$$GS1 = \frac{\left(\sum_{i=1}^{k} (d_i - \bar{d}) Y_{i\cdot}\right)^2}{\sum_{i=1}^{k} (d_i - \bar{d})^2 \sum_{j=1}^{m_i} (Y_{ij} - \bar{Y})^2}.$$

Table 3

Trend patterns of mean response from five dose groups used in generating data for simulation

No.	μ_1	μ_2	μ_3	μ_4	μ_5	Pattern
1	2	2	2	2	2	Uniform
2	2.0	2.3	2.9	4.0	6.0	Monotone
3	2.0	3.8	5.5	6.0	5.0	Umbrella

We call the test based on this statistic the GS1 test. The GS1 test is identical to the special case of the trend test for clustered binary data discussed in Boos (1992), Carr and Gorelick (1995), and Lefkopoulou, Rotnitzky, and Ryan (1996) when there is only one subject in each cluster. Note that the numerator of the GS1 is equivalent to that of the QL test by Piegorsch and Bailer (1997). Their empirical variance estimator is $\Sigma_{i=1}^{k} (d_i - \bar{d})^2 \Sigma_{j=1}^{m_i} (Y_{ij} - \bar{Y}_{i\cdot})^2$.

When $r = 2$, the GSr is written as

$$GS2 = \left(\sum_i a_{1i} Y_{i\cdot}, \ \sum_i a_{2i} Y_{i\cdot}\right) \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix}^{-1} \times \begin{pmatrix} \sum_i a_{1i} Y_{i\cdot} \\ \sum_i a_{2i} Y_{i\cdot} \end{pmatrix},$$

where $v_{sl} = \Sigma_i a_{si} a_{li} \Sigma_j (Y_{ij} - \bar{Y})^2$. We call the test based on this statistic the GS2 test. It will be shown below by simulation that the GS2 test has high power in detecting umbrella trends. The GSr test assumes no specific distribution to represent extra-Poisson variability. In the next section, however, its behavior is examined under negative binomial distribution. Specifically, assume that $\{Y_{ij}\}_{j=1,\ldots,m_i}$ are independent and

$$P(Y_{ij} = y) = \frac{\Gamma\left[y + \frac{1}{\varphi}\right]}{y!\Gamma\left(\frac{1}{\varphi}\right)} \left(\frac{\varphi\mu_i}{1 + \varphi\mu_i}\right)^y \frac{1}{(1 + \varphi\mu_i)^{1/\varphi}},$$

giving $\mathrm{E}(Y_{ij}) = \mu_i$ and $\mathrm{var}(Y_{ij}) = \mu_i + \varphi\mu_i^2$, where $\varphi > 0$ is the dispersion parameter.

3. Numerical Evaluation

The response patterns of numbers 1, 2, and 3 in Table 3 are referred to as the uniform, monotone, and umbrella patterns. In addition to the GS1, QL, and GS2, the Cochran and Armitage test (CA test) (Cochran, 1954; Armitage, 1955) is examined for each response pattern, with means shown in Table 3. The dose levels are set at 1, 2, 3, 4, and 5, and the nominal test level is set at 0.05. Ten thousand data are generated from each distribution. The number of observations (m_i) per dose group is the same for all dose groups, and its value is set at $3, 4, \ldots, 20$.

Empirical Type I errors are shown in Figure 1 for the uniform response. Panel (a) exhibits the errors for the Poisson distribution with no extra-Poisson variability. The panel shows that the Type I errors of the tests are quite close to the nominal level except for the QL test at small m. Panels (b), (c), and (d) exhibit the errors when the underlying distribution is the negative binomial with $\varphi = 0.1$, 0.2, and

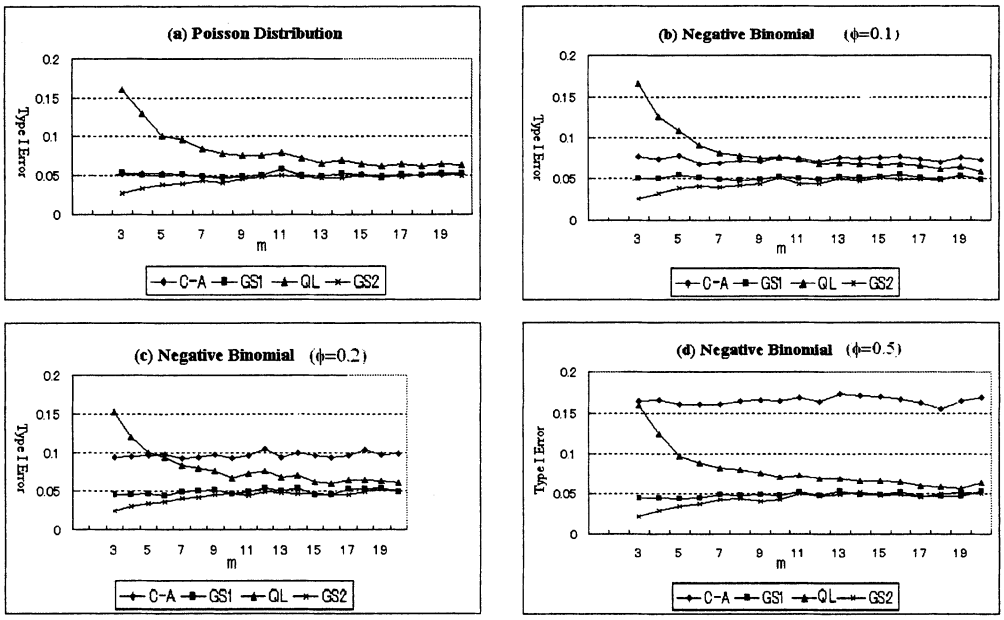

Figure 1. Empirical type I error for the uniform response.

0.5, respectively. The panels show that the Type I errors of the CA test deviate substantially from the nominal level, confirming the previous findings (Margolin et al., 1981; Boos, 1993; Carr and Gorelick, 1965), and that the Type I errors of the QL test are inflated, in particular for small m. Note that the QL test is an asymptotic test when $m \to \infty$. In contrast, the GS1 and GS2 tests perform reasonably well by keeping the Type I error close to 0.05, even for small m.

Next we examine the power of the tests. Empirical powers of the CA, GS1, QL, and GS2 tests are considered for the Poisson distribution, but the CA test is omitted from consideration of the negative binomial distribution because that test violates the nominal size substantially. The power of the QL test is included throughout but is omitted in the comparison below when m is small for the same reason as the CA test. Figure 2 displays the power of the tests for the monotone response. Panel (a), for the Poisson, shows that the power of the CA test is the highest, as expected. Panels (b), (c), and (d), for the negative binomials distributions with the same dispersion parameters as in Figure 1, show that the GS1 test has higher power than the GS2 test throughout, but the power of the GS1, QL, and GS2 tests begin to converge as m increases.

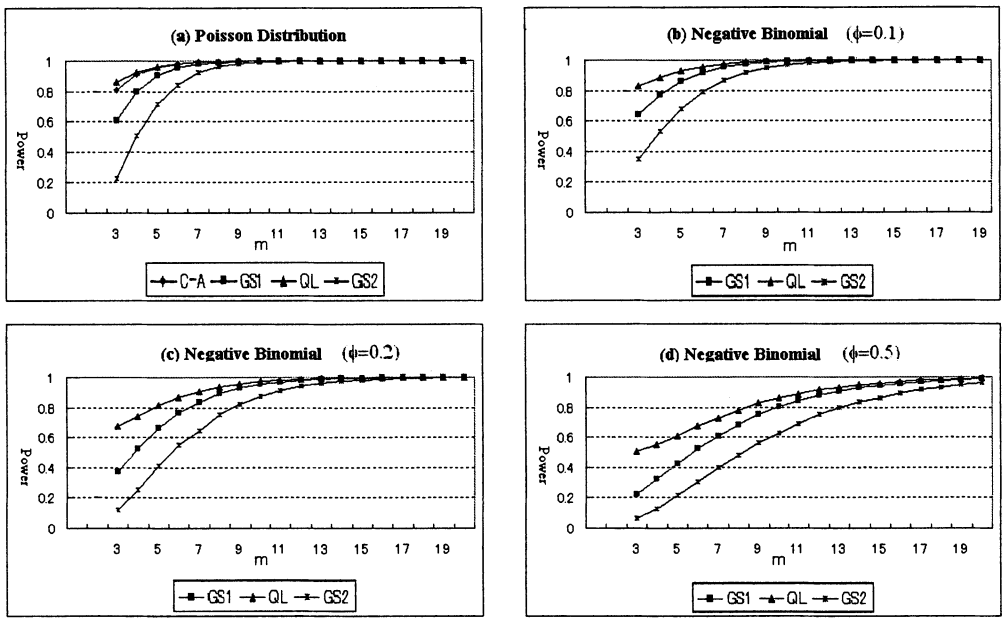

Figure 2. Empirical power for the increasing monotone response.

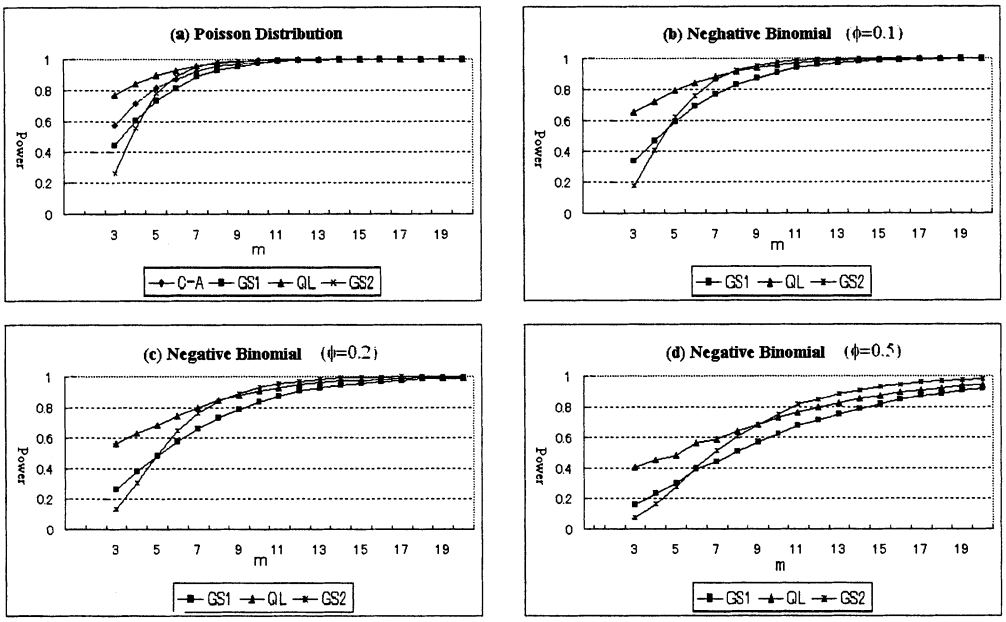

Figure 3. Empirical power for the umbrella response.

Figure 3 shows the power of the tests for the umbrella response. Again, panel (a) is for the Poisson distribution without the extra-Poisson variability. The panel shows that the CA test is still superior to the others and that the powers of the all four tests converge when m is large. Panels (b), (c), and (d) are for the same negative binomial distributions as the previous figures. The panels show that, when m is small, the GS1 test has a little higher power than the GS2 test, but otherwise, the power of the GS2 test is the highest and the power of the GS2 test decreases slightly as the value of the dispersion parameter increases.

4. Applications

4.1 C. dubia *Data*

The values of CA, GS1, QL, and GS2 computed from Table 1 are 182.655, 25.306, 304.252, and 25.342, respectively. Since the response is monotone, the results of the simulation in the previous section suggest that the GS1 test is preferable. The GS1 test indicates a significant result with p-value less than 0.01, showing nitrofen induces strongly a significant downward trend in *C. dubia* data.

4.2 Salmonella *Data*

The results of the Fisher test (Fisher, 1950) for overdispersion are given in the second column of Table 4. They indicate the presence of overdispersion in replicate 1 and the total data. Columns 3, 4, 5, and 6 in Table 4 list the p-values of the CA, GS1, QL, and GS2 tests. Since $m = 3$ in each replicate, the QL test is only applied to the total data for the reason given above. Also, for the same reason, the CA test is not applied to the data in replicates 1 and 3. Recall that the dose–response in each replicate increases over low doses but then has a down turn in higher doses. The previous simulations indicate that the GS2 test is preferable in this case. The table shows that the CA, GS1, and QL tests all fail to detect the umbrella response, whereas it is detected by the GS2 test. The Jonckheere–Terpstra-type recursive test (Simpson and Margolin, 1986) provides upper bounds of the p-values; those upper bounds for replicates 1, 2, and 3, respectively, are 0.016, 0.015, and 0.01.

5. Discussion

The GS1 and GS2 tests are proposed for testing a trend under extra-Poisson variability, and their behaviors are compared with the CA and QL tests. It is shown by simulation (i) that, if extra-Poisson variability exists, the CA test looses its validity but the GS1 test and GS2 test do not; (ii) that the GS1 test is superior to the GS2 test in detecting monotone response and the GS2 test is superior to the GS1 test in detecting an umbrella response unless m is very small; and (iii) that the QL test overstates the Type I errors and should not be used unless m is large. Note that the denominator of QL is smaller than that of GS1 and thus QL is always larger than GS1; in particular, QL=∞ when m=1. This would account for the inflation of the Type I error of the QL test. The GSr test is also an asymptotic test for large m, but it employs pooled empirical variance and is more faithful to the Type I errors than the QL test even when m is small.

Table 4
P-value for overdispersion test and trend tests

Data	Fisher test for overdispersion	Trend test			
		CA	GS1	QL	GS2
Replicate 1	**0.001**	—	0.882	—	0.003
Replicate 2	0.363	0.125	0.399	—	0.048
Replicate 3	0.794	0.306	0.534	—	0.019
Total data	**0.000**	—	0.442	0.120	**0.001**

Acknowledgements

We are grateful to two anonymous referees for their comments, which helped improve greatly the articles.

Résumé

Des tests de tendances permettant de tester la monotonie de tendances ou leur forme en ombrelle (passant d'une monotonie croissante à une monotonie décroissante) dans des données de dénombrement sont proposés quand les données montrent une variabilité extra-Poissonnienne. Les tests proposés, que l'on appelle test GS1 et test GS2, sont construits en appliquant un vecteur de scores orthonormal à un test de scores généralisé sous un modèle log-linéaire de r-ème ordre. Ces tests sont comparés par simulation au test de Cochran-Armitage ainsi qu'au test de quasi-vraisemblance de Piegorsch et Bailer (1997). On montre que le test de Cochran-Armitage ne devrait pas être utilisé en présence d'une variabilité extra-Poissonienne; que pour détecter une tendance monotone le test GS1 est supérieur aux autres; et que le test GS2 a une grande puissance pour détecter les tendances en ombrelle.

References

Armitage, P. (1955). Test for linear trend in proportions and frequencies. *Biometrics* **11,** 375–386.

Bailer, A. J. and Oris, J. T. (1993). Modeling reproductive toxicity in *Ceriodaphnia* tests. *Environmental Toxicology and Chemistry* **12,** 787–791.

Boos, D. D. (1992). On generalized score test. *The American Statistician* **46,** 327–333.

Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. *Applied Statistics* **33,** 38–44.

Breslow, N. E. (1990). Test of hypotheses in overdispersed Poisson regression and other quasi-likelihood models. *Journal of the American Statistical Association* **85,** 565–571.

Carr, G. J. and Gorelick, N. J. (1995). Statistical design and analysis of mutation studies in transgenic mice. *Environmental and Molecular Mutagenesis* **25,** 246–255.

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics* **10,** 417–451.

Fisher, R. A. (1950). The significance of deviations from expectation in a Poisson series. *Biometrics* **6,** 17–24.

Lefkopoulou, M., Rotnitzky, A., and Ryan, L. (1996). Trend tests for clustered data. In *Statistics for Toxicology,* B. J. T. Morgan (ed), 179–197. Oxford: Clarendon Press.

Margolin, B. H., Kaplan, N., and Zeiger, E. (1981). Statistical analysis of the Ames salmonella microsome test. *Proceedings of the National Academy of Sciences* **76,** 3779–3783.

Neuhäuser, M., Seidel, D., Hothorn, L. A., and Urfer, W. (2000). Robust trend tests with application to toxicology. *Environmental and Ecological Statistics* **7,** 43–56.

Piegorsch, W. W. and Bailer, A. J. (1997). *Statistics for Environmental Biology and Toxicology.* London: Chapman and Hall.

Simpson, D. G. and Margolin, B. H. (1986). Recursive nonparametric testing for dose–response relationship subject to down turns at high doses. *Biometrika* **73,** 586–596.

Williams, D. A. (1982). Extra-binomial variation in logistic linear models. *Applied Statistics* **31,** 144–148.

Received April 2001. *Revised January* 2002.
Accepted January 2002.

Appendix

Let $\ell(\beta)$ be the log-likelihood function obtained by assuming the Poisson distribution for the data incorporated with the log-linear model (1). Put

$$S(\boldsymbol{\beta}) = \left(\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_t}\right)_{(r+1)\times 1},$$

$$I_Y(\boldsymbol{\beta}) = \left(-\frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_t \partial \beta_u}\right)_{(r+1)\times(r+1)},$$

$$D_Y(\boldsymbol{\beta}) = \left(\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_t}\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_u}\right)_{(r+1)\times(r+1)},$$

and let $S(\beta)' = (S'_{(1)}, S'_{(2)})$, where $S_{(1)}$ is 1×1 and $S_{(2)}$ is $r \times r$. The matrices above are partitioned accordingly, e.g., $I_{Y(11)}$ is 1×1, $I_{Y(12)}$ is $1 \times r$, and so on. Evaluating those matrices at $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$, where $\tilde{\boldsymbol{\beta}}$ is the restricted maximum likelihood estimator of $\boldsymbol{\beta}$ under H_0^*: $\boldsymbol{\beta}_{(2)} = \mathbf{0}$, Boos (1992) defined the generalized score test for testing H_0^* against H_1^*: $\boldsymbol{\beta}_{(2)} \neq \mathbf{0}$ as $T_{\mathrm{GS}} = \tilde{S}'_{(2)} \tilde{\mathbf{V}}(\tilde{S}_{(2)})^{-1} \tilde{S}_{(2)}$, where

$$\begin{aligned}\tilde{\mathbf{V}}(\tilde{S}_{(2)}) = {} & \tilde{D}_{Y(22)} - \tilde{I}_{Y(21)} \tilde{I}^{-1}_{Y(11)} \tilde{D}'_{Y(21)} \\ & - \tilde{D}_{Y(21)} \tilde{I}^{-1}_{Y(11)} \tilde{I}'_{Y(21)} \\ & + \tilde{I}_{Y(21)} \tilde{I}^{-1}_{Y(11)} \tilde{D}_{Y(11)} \tilde{I}^{-1}_{Y(11)} \tilde{I}'_{Y(21)}.\end{aligned}$$

Using orthonormal properties of the score vectors, we have $\tilde{I}_{Y(21)} = \tilde{I}'_{Y(12)} = \mathbf{0}$ and the variance estimate $\tilde{\mathbf{V}}(\tilde{S}_{(2)})$ takes the simple form of $\tilde{D}_{Y(22)}$ given in the text.