Environmental Health Perspectives Supplements
Vol. 102 Suppl 1:77-82 1994

Statistical Models for Genetic Susceptibility in Toxicological and Epidemiological Investigations

by Walter W. Piegorsch

Models are presented for use in assessing genetic susceptibility to cancer (or other diseases) with animal or human data. Observations are assumed to be in the form of proportions, hence a binomial sampling distribution is considered. Generalized linear models are employed to model the response as a function of the genetic component; these include logistic and complementary log forms. Susceptibility is measured via odds ratios of response, relative to a background genetic group. Significance tests and confidence intervals for these odds ratios are based on maximum likelihood estimates of the regression parameters. Additional consideration is given to the problem of gene-environment interactions and to testing whether certain genetic identifiers/categories may be collapsed into a smaller set of categories. The collapsibility hypothesis provides an example of a mechanistic context wherein nonhierarchical models for the linear predictor can sometimes make sense.

Introduction

Recent technological advances in biomedical experimentation have greatly improved identification of genetic damage (*1*) and recognition of genetic factors that may affect disease susceptibilities in animal (*2*) and human (*3*) subjects. Molecular genetic techniques are now used to idenitify specific genotypes or genetic patterns in individuals affected by some disease or, for example, exhibiting cancer. For instance, the role of genetic factors in lung tumor onset and progression has been recently highlighted (*4–7*), as have genetic components in the development of human bladder tumors and other cancers (*8–10*). To study these effects, various biochemical, cytogenetic, and molecular probes are used (*11–14*), and epidemiologic research has moved to use these methods in studies of disease/cancer susceptibility (*15,16*). Of interest is whether individuals in various genetic categories display greater risk of cancer or disease than those identified in some background, control, or genetic "wild-type" category. Statistical models and methods for assessing these risks in both animal experiments and human population studies are described in the following sections.

Statstical Models: Generalized Linear Forms

Assume the existence of $T \geq 2$ genetic susceptibility groups or categories, identified without error via some form of biological/biomolecular probe, and indexed by $i=0,\ldots,T-1$. For instance, the experimental study design might compare the effects of a set of T different genotypes or polymorphisms on the cancer or disease under study. The group at $i = 0$ is considered the background group to which susceptibility comparisons are to be made. From each prospectively sampled group, a count, Y_i, of individuals exhibiting tumors is recorded. This is compared to the total number, N_i, of individuals in each group ($i = 0,\ldots T-1$). For example, one might compare the proportion of mice developing tumors between two different inbred strains. The Y_is are assumed to take the binomial distribution (*17*) with (known) sample size parameter N_i and probability parameter p_i.

As noted above, susceptibility in the ith genetic group is measured relative to the response in the background, $i = 0$ group. This is quantified via the odds ratios

$$\psi_i = \frac{p_i}{1-p_i}\frac{1-p_0}{p_0}$$

To model the p_i and the ψ_i, it is natural to consider the logistic form (*18*)

$$log\left\{\frac{p_i}{1-p_i}\right\} = \mu + \alpha_i \qquad (1)$$

(i=0,...T−1). This is a generalized linear model (*19*) that links a function (here, the logit) of the response probability to the linear predictor $\mu + \alpha_i$. (In this over-saturated form, the model requires an identifiability constraint: $\alpha_0=0$.)

Under this logistic model, the odds ratios take the simple form $\psi_i = exp\{\alpha_i\}$, $\forall i$. Estimation and testing of regression parameters associated with these odds ratios is performed with computer programs and packages such as SAS (*20*) or GLIM (*21*) that fit

Statistics and Biomathematics Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709.

Present address: Dept. of Biostatistics, University of South Carolina, Columbia, SC 29208

This paper was presented at the International Biostatistics Conference on the Study of Toxicology that was held May 13–25, 1991, in Tokyo, Japan.

the logistic model and provide maximum likelihood (ML) estimates, likelihhod ratios (LR), etc.

The null effect specifies constant susceptibility relative to the background group; this corresponds to equality of odds ratios:

$$H_0: \psi_1 = \psi_2 = \cdots = \psi_{T-1} = 1 \ .$$

Under the logistic model, this becomes

$$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_{T-1} = 0 \ .$$

Departures from H_0 suggest susceptibility somewhere among the T groups. Of typical interest is identification of increased cancer susceptibility; this is indicated for the ith group when $\psi_i > 1$, which occurs under the logistic model if and only if $\alpha_i > 0$. Thus, tests of H_0 against one-sided alternatives are readily available by testing the sign of α_i.

To test H_0 against a global one-sided departure, H_1: $\alpha_i > 0, \forall i$, (i.e, that all odds ratios exceed one) one appeals to the large-sample normality of the ML estimate $\hat{\alpha}_i$. A recommended approach that identifies simultaneously individual departures from H_0 (i.e, which of the individual α_i are positive) is based on a modification of the well-known Bonferroni inequality (*22*). Begin with the individual Wald test (*23*) p-values

$$P_i = 1 - \Phi\{ \hat{\alpha}_i / se(\hat{\alpha}_i) \} ,$$

where $se(\hat{\alpha}_i)$ is the large-sample standard error of $\hat{\alpha}_i$, and $\Phi(\bullet)$ is the cumulative distribution function from a standard normal distribution. Order these values from smallest to largest; denote $P_{(i)}$ as the ith smallest ordered probability. Set the desired simultaneous confidence level to $1-a$. Then, calculate the index, $\mathcal{M}$, which is the largest i such that

$$P_{(T-1-i+v)} > \frac{va}{i}$$

(this inequality must hold for every value of $v=1,\ldots,i$, at each i under scrutiny). Conclude that α_i is significantly greater than 0 with simultaneous confidence $1-a$, if $P_i \leq a/\mathcal{M}$. (If $\mathcal{M}$ cannot be calcualted, conclude that α_i is significantly greater than 0 at all levels of i). Other aspects of this and related approaches to one-sided testing are described in detail elsewhere (*24*).

If a quantification via some intensity variable, say, v_i, exists for each genetic category/group, the logistic model may be enhanced by incorporating this quantitative information. A simple (logistic) linear model is

$$log \left\{ \frac{p_i}{1-p_i} \right\} = \mu + \theta v_i$$

($i=0,\ldots,T-1$), where it is assumed that $v_0 < v_1 < \ldots < v_{T-1}$. Under this dose-response model, the odds ratios are

$$\psi_i = \theta (v_i - v_0) , \ \forall i .$$

Hence, $\psi_i = 0$ (for all i) if and only if $\theta = 0$, while $\psi_i > 1$ (for all i) if and only if $\theta > 0$. One-sided testing is again of interest, although it takes on a simpler formulation in the dose–response setting, since only one parameter, θ, is assessed. For example, the Wald test of $H_0: \theta = 0$ versus $H_1: \theta > 0$ rejects H_0 when $\hat{\theta} / se(\hat{\theta}) > z_a$, where z_a, is the $1-a$ quantile from a standard normal distribution.

Table 1. Lung tumor susceptibility data.

	Group	
	0.70-kb homozygous (i=0)	0.70-kb/0.55-kb heterozygous (i=1)
lung tumors:	8/16	11/12

Example 1

To illustrate use of the logistic model in Equation 1, consider the lung tumor susceptibility data given by Ryan et al. (*4*). These authors considered susceptibility to the known murine carcinogen urethan by examining specific allelic forms of the *K-ras-2* proto-oncogene in recombinant offspring from crosses of inbred strains of mice. The susceptibility allele is characterized by a shorter initial exon (length 0.55 kb) compared to the normal allele (length 0.70 kb). The mice under study were known to be either homozygous for the 0.70-kb allele, or heterozygous (0.70 kb/0.55 kb). If the 0.55-kb allele were to confer or otherwise indicate increased susceptibility to lung tumorigenesis, heterozygous mice would exhibit greater lung tumor rates and thus an odds ratio relative to the homozygous mice greater than one.

The data for the T=2 groups are shown in Table 1. Applying a logistic model to these data gives an ML estimate of the regression parameter as $\hat{\alpha}_1 = 2.398$, with $se(\hat{\alpha}_1) = 1.16$. A Wald test of the no susceptibility hypothesis $H_0: \alpha_1 = 0$ yields a test statistic of

$$Z = \frac{\hat{\alpha}_1}{se(\hat{\alpha}_1)} = 2.07,$$

with one-sided p-value equal to 0.019. From the ML estimate of α_1, one finds the ML estimate of the odds ratio to be $exp[\hat{\alpha}_1] = \hat{\psi}_1 = 11.0$. Large-sample 95% confidence limits for ψ_1 are given by

$$exp\{ \hat{\alpha}_1 \pm z_{a/2} se(\hat{\alpha}_1) \} .$$

For these data, this yields $1.14 < \psi_1 < 106.45$.

Based on this analysis, it seems fair to conclude that the 0.70-kb/0.55-kb heterozygous genotype exhibited moderately increased risk of murine lung tumorigenesis relative to the homozygous genotype. Sample sizes are small; total samples of at least 100 have been suggested to achieve nominal operating characteristics in one-sided testing under the logistic model (*24*). Thus, further experimentation and analysis are required before unequivocal conclusions can be reached as to the heterozygote susceptibility in this setting.

Two-Way Models and Gene–Environment Interactions

In example 1, only T=2 genetic groups were examined for murine lung cancer susceptibility. With T=2 groups, a number of possible analyses can identify susceptibility, including 2 × 2 contingency table calculations with χ^2 tests (*25*), Fisher exact

tests (*26,27*), etc. These approaches often provide similar inferences. For example, the one-sided p-value from the Fisher exact test comparing the two proportions in example 1 is 0.024, almost identical with the value of 0.019 achieved with the Wald test of α_1.

The usefulness of the logistic model is more evident, however, in cases with many genetic categories or a dose–response under study, or when additonal factors or other sources of variability are identified and examined as part of the susceptibility analysis. For example, with human lung cancer, susceptibility may associate with genetic effects, lifestyle-related factors such as cigarette smoking, occupational or environmental exposures to pulmonary genotoxins, or a combination of these factors (*15*). (An application of the logistic model to such data is presented in example 2, below.) Indeed, recognition is growing that genetic susceptibility must be studied in the context of external environmental exposures that might initiate or contribute to disease progression (*28*). In these cases, the linear predictor in Equation 1 is easily extended, and the logistic model facilitates estimation of the additional parameters.

Consider the case of two factors: the genetic factor continues to be indexed by $i=0,\ldots,T-1$, and an additional, "environmental" factor is now considered, indexed by $j=0,\ldots,J-1$. The associated two-way extension of Equation 1 takes the $T \times J$ form

$$log\left\{\frac{p_{ij}}{1-p_{ij}}\right\} = \mu + \alpha_i + \beta_j + \gamma_{ij} \qquad (2)$$

($i=0,\ldots,T-1$; $j=0,\ldots,J-1$). As above, for reasons of estimability one assumes that

$$\alpha_0 = \beta_0 = \gamma_{i0} = \gamma_{0j} = 0\,, \quad \forall i,j\,.$$

The interaction parameters in Equation 2, γ_{ij}, relate to the odds ratios via

$$\gamma_{ij} = ln\{\psi_{ij}/\psi_{i0}\psi_{0j}\}\,,$$

where

$$\psi_{ij} = \frac{p_{ij}(1-p_{00})}{p_{00}(1-p_{ij})}.$$

Thus, H_0:$\gamma_{ij}=0$ is equivalent to a simple multiplicative relationship among odds ratios (*29–31*):

$$\psi_{ij} = \psi_{i0}\psi_{0j}\,.$$

Multiple-level multiplicative interactions are assessed by simultaneously identifying equality to 0 for each γ_{ij} of interest. The $(T-1)(J-1)$ degree of freedom (df) hypothesis H_0:$\gamma_{ij}=0$ ($\forall\, i,j$) is equivalent to H_0:$\psi_{ij}=\psi_{i0}\psi_{0j}$, ($\forall\, i,j$).

In those instances where no departure from multiplicative interaction is evidenced, it may be of interest to estimate various parameters from the reduced model. For instance, under $\psi_{ij}=\psi_{i0}\psi_{0j}$, the ML estimate of ψ_{ij} is simply

$$\hat{\psi}^0_{ij} = exp\{\hat{\alpha}_i + \hat{\beta}_j\}\,.$$

Large-sample $1-a$ confidence limits for ψ_{ij} are given by

$$exp\{\hat{\alpha}_i + \hat{\beta}_j \pm z_{a/2}\,[se^2(\hat{\alpha}_i) + se^2(\hat{\beta}_j) + 2\,cov(\hat{\alpha}_i,\hat{\beta}_j)]^{1/2}\}\,,$$

where $\text{cov}(\hat{\alpha}_i,\hat{\beta}_j)$ is the estimated covariance between $\hat{\alpha}$ and $\hat{\beta}_j$. (If either index is 0 the corresponding standard error will be 0 since we set $\alpha_0=\beta_0=0$. Similarly, $\text{cov}(\hat{\alpha}_i,\hat{\beta})=0$ if $i=0$ or $j=0$.)

Collapsibility

Collapsibility over genetic categories is another potential area of interest in studies of genetic susceptibility to disease. One questions whether the different genotypes have equivalent effects and may be collapsed into one or a small group of categories. For example, a single-locus, two-allele (say, B and b) system generates three genetic categories: BB,Bb,bb. If B behaves as a simple dominant allele, we can collapse these three categories into two: B– and bb. It may be of interest to assess any such collapse statistically.

Obviously, the nature of the collapsibility hypothesis will depend on the gene under study. For instance, if one encountered a situation where the genetic factor is important to the detoxification of an environmental exposure, or is important in either producing or inactivating a toxic metabolic product of the exposure, then there may be no apparent genetic effect in those individuals without the exposure. In such a case, this may indicate collapsibility at only certain levels of the exposure variable. The logistic model provides a means for assessing collapsibility of this (or any) sort in the $T \times J$ setting. At any fixed environmental level, say $j = j_0$, consider collapsibility over C of the genetic levels, indexed by $i_1,\ldots,i_C$. This is expressed as

$$H_0\colon \psi_{i_1j_0} = \psi_{i_2j_0} = \cdots = \psi_{i_Cj_0}\,.$$

Under the logistic model, this corresponds to

$$H_0\colon \alpha_{i_1} + \gamma_{i_1j_0} = \alpha_{i_2} + \gamma_{i_2j_0} = \cdots = \alpha_{i_C} + \gamma_{i_Cj_0}\,.$$

Departure from H_0 is assessed via a generalized LR statistic (*32*), for example, with a limiting χ^2 distribution on $C-1$ df. Extensions to collapsibility over multiple levels of j are straightforward.

Notice that at $j=0$ (i.e., at the "background" environmental level) the collapsibility hypothesis is

$$H_0\colon \alpha_{i_1} + \gamma_{i_10} = \alpha_{i_2} + \gamma_{i_20} = \cdots = \alpha_{i_C} + \gamma_{i_C0}\,.$$

Under the identifiability constraints $\gamma_{i0}=0$, for all i, this simplifies to

$$H_0\colon \alpha_{i_1} = \alpha_{i_2} = \cdots = \alpha_{i_C}\,.$$

Table 2. Case-control data on debrisoquine metabolism for lung cancer.

	Metabolization category		
	PM	IM	EM
Cases	3	11	116
Controls	9	52	81

Abbreviations: PM, poor metabolizer; IM, intermediate metabolizer, EM, extensive metabolizer.

Thus, collapsibility at the environmental background corresponds to collapsibility among genetic main effects, as would be expected. This illustrates a situation where a nonhierarchial model (with interactions, but not all main effects, fully modeled) makes sense from a mechanistic perspective, a paradigm not typically encountered in (generalized) linear modeling.

Collapsibility can also be assessed under a no-interaction condition, i.e, when $\gamma_{ij}=0\ \forall\ i,j$. The collapsibiltiy hypothesis becomes

$$H_0:\ \alpha_{i_1} = \alpha_{i_2} = \cdots = \alpha_{i_C}\ ,$$

at any j. This condition is independent of j, however, so collapsibility at any j when $\gamma_{ij}=0\ \forall i,j$ corresponds to collapsibility over all j. Joint collapsibility follows similarly. Suppose collapsibility is considered over the two nonoverlapping categories indexed by $i_1,\ldots,i_C$ and $I_1,..,I_D$. Then, joint collapsibiltiy is expressed as a $(C-1)+(D-1)$ df hypothesis:

$$H_0: \psi_{i_1} = \psi_{i_2} = \cdots = \psi_{i_C}\ ;\ \psi_{I_1} = \psi_{I_2} = \cdots = \psi_{I_D}\ ,$$

corresponding to

$$H_0: \alpha_{i_1} = \alpha_{i_2} = \cdots = \alpha_{i_C}\ ;\ \alpha_{I_1} = \alpha_{I_2} = \cdots = \alpha_{I_D}\ .$$

Of course, one should have some *a priori* basis for considering certain sets of genetic types as reasonable candidates for collapse because exploratory analyses over all possible collapsings run the risk of data overinterpretation.

Applications in Epidemiology

Once a potential susceptibility gene has been identified in humans, its association with disease can be tested in epidemiologic population studies using case-control study designs (*33*). Sampling in a case-control study is carried out separately for cases and controls in a retrospective manner and thus in effect is conditioned on disease status (*34*). As is well known, however, one can reverse this conditioning and model the logit of the risk of disease as a function of covariates as in Equation 1 or 2, treating the data as if they had arisen prospectively. The resulting regression coefficients are asymptotically unbiased for the associated log odds ratios (*35,36*), and likelihood ratio testing based on the prospective logistic model is valid when applied to case-control data (*37*). Thus the (prospective) logistic models descibed above are applicable in retrospective (and prospective) epidemiologic studies where genetic susceptibility is under examination.

Table 3. Interaction of genetic susceptibility evidenced by debrisoquine phenotype and asbestos exposure.

	Cases		Controls	
Asbestos exposure	PM/IM	EM	PM/IM	EM
−	14	97	53	68
+	3	47	15	17

Abbreviations: PM, poor metabolizer; IM, intermediate metabolizer; EM, extensive metabolizer

Example 2

In a case-control study, Caparaso et al. (*38*) examined individual subjects' abilities to metabolize the drug debrisoquine and related these metabolic activities to lung cancer susceptibility. Increased ability to metabolize agents such as debrisoquine is conjectured to indicate increased cancer susceptibility because heterogenity in drug metabolism may associate with heterogeneity in cancer susceptibility. (For example, the drug metabolism pathway may play a role in carcinogenesis metabolism, either by deactivating a carcinogen or by activating a proto-carcinogen into a carcinogen.) Debrisoquine metabolism is polymorphic in humans: most individuals receiving the drug rapidly excrete large amounts of debrisoquine metabolite (*39*); these individuals are "extensive metabolizers" (EM). Some individuals excrete reduced amounts of the metabolite or excrete the drug almost unchanged. They are "intermediate" (IM) or "poor" metabolizers (PM), respectively (*5*). These polymorphisms lead to $T=3$ genetic categories for study.

An initial question of interest is whether the intermediate or poor metabolizers truly constitute two distinct genetic classes with respect to their lung cancer susceptibility. That is, is collapsibility evidenced between PM and IM? To study this question, Caporaso et al. (*38*) reported the case-control data shown in Table 2. Since the logistic model is applicable to this retrospective sampling scenario (*37*), we consider the prospective form in Equation 1 to model the genetic effects. As suggested above, collapsibility of PM and IM categories corresponds to equality of main effects parameters: $H_0:\alpha_1=\alpha_2$. To test this hypothesis, a GLIM (*21*) analysis of these data yields a 1 df LR statistic of 0.178. No significant departure is evidenced, and we conclude that the data support the contention that PM and IM metabolizers exhibit similar susceptibilities to lung cancer.

Caporaso et al. (*38*) also reported data on the potential interaction of genetic susceptibility (as evidenced by debrisoquine phenotype) and environmental factors, such as asbestos exposure (Table 3). Notice that the data reflect the previous recognition of PM/IM collapsibility. Referring now to the two-way model from Equation 2, no interaction between debrisoquine phenotype and asbestos exposure corresponds to testing H_0: $\gamma_{11}=0$. The LR statistic for this significance test is 1.609, on 1 df. The corresponding P-value is 0.205. No evidence is seen for a significant interaction between the genetic and environmnetal factors.

Data Truncation

In some settings, the experimental end point may involve the number of occurrences of some phenomenon, such as the number of tumors seen in a certain organ of an experimental animal (*40*) or the number of cells in a tissue or culture responding to a chemical stimulus (*41*). Denote the random variable

associated with this discrete-valued response by U. If the observing mechanism or technique is such that only the occurrence of a non-null state is recorded (e.g., "no tumors" versus "some tumors"), the data will be truncated into a dichotomous response. The observed variable becomes

$$Y = \begin{cases} 1 & \text{if } U > 0 \\ 0 & \text{if } U = 0 \end{cases},$$

a binary variate with probability of response $p=Pr[U>0]$. Sobel and Elashoff (*42*) have referred to this sampling scheme as (binomial) "group testing"; also see Chen and Swallow (*43*). When interest centers on the nonresponse, $Pr[Y=0]$, the data are often referred to as "Hansen frequencies" (*44*), based on applications of E. W. Hansen's work in the behavioral sciences (*45*).

This sort of data truncation could occur in a susceptibility study, where multiple tumors occur in each individual, but only the presence or absence of the cancer is noted. Thus, for the kth individual in the ith genetic category or group, one observes Y_{ik} as the indicator of individual tumorigenic response. In this truncation scenario one often takes the U_{ik} as independently distributed Poisson random variates, with per-group means following the one-way model $\mu+\alpha_i$ ($i=0,\ldots,T-1$). Thus a new generalized linear model for p_i is induced:

$$p_i = \Pr[U_{ij} > 0] = 1 - \Pr[U_{ij} = 0] = 1 - exp\{-\mu - \alpha_i\}.$$

The result is a complementary log regression equation:

$$-log\{1 - p_i\} = \mu + \alpha_i . \qquad (3)$$

This construction, based on Poisson occurrence rates, was discussed in detail by Cochran (*46*), who had in mind application to bacterial concentrations in suspension and the planning of dilution experiments. He suggested that the concept was fairly well known, starting with the work of McCrady (*47*) on the concentration of organiams in liquids.

In general, if a set of explanatory variables, $x_1,\ldots,x_E$, are associated with Y, a generalized linear model could be fit under this data truncation using the complementary log regression with linear predictor

$$\omega_0 + \omega_1 x_1 + \cdots + \omega_E x_E .$$

The ω parameters are unknown regression coefficients, fit via maximum likelihood.

Odds ratios under the complementary log model are somewhat different from the simple forms encountered under the logistic model in Equation 1. For the simple one-way model from Equation 3, one has

$$\psi_i = \frac{exp\{\alpha_i\} - e^{-\mu}}{1 - e^{-\mu}}$$

($i=1,\ldots,T-1$). Point estimates and large-sample standard errors under Equation 3 are still available for ψ_i, using maximum likelihood. Also, the null effect model

$$H_0: \psi_1 = \psi_2 = \cdots = \psi_{T-1} = 1$$

again corresponds to

$$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_{T-1} = 0 ;$$

departures from H_0 continue to suggest susceptibility among the T groups.

The complementary log model shares with the logistic model the characteristic that increased cancer susceptibility, evidenced when $\psi_i>1$, occurs if and only if $\alpha_i>0$. Thus, tests of the individual null susceptibility hypothesis $H_0:\psi_i=1$ against the one-sided alternative $H_1:\psi_i>1$ are again available by testing the sign of α_i, in similar fashion to the logistic model. Hypothesis testing extensions to global, one-sided departures are also available (*24*).

Example 1 *(continued)*

To illustrate use of the complementary log model from Equation 3, consider again the K-*ras*-2/lung tumor susceptibility data described earlier. For the inbred recombinant mice studied in that experiment, it is common to observe multiple lung neoplasms per mouse (*4*). Reporting the data as dichotomous outcomes therefore involves a $U \rightarrow Y$ data truncation of the form considered herein. The complementary log formulation becomes a viable model candidate for quantitative assessment of the cancer susceptibility.

Applying a complementary log model to these data gives an ML estimate of the regression parameter as $\hat{\alpha}_1=1.792$, with $se(\hat{\alpha}_1) = 0.9895$. A Wald test of the no-susceptibility hypothesis $H_0:\alpha_1=0$ yields a test statistic of $Z = 1.81$, with one-sided p-value equal to 0.035. Again, increased susceptibility is evidenced, although with slightly lesser significance than that exhibited with the logistic analysis. (The call for larger sample sizes remains valid, and is perhaps in greater evidence here.) Additional similarity is seen with the ML estimate of the odds ratio, constructed from the ML estimate of α_1. Under the complementary log formulation, $\hat{\psi}_1 = 11.003$ for these data, an almost indistinguishable change from the logistic estimate reported above.

Thanks are due to Joseph K. Haseman, David G. Hoel, Jack A. Taylor, Clarice R. Weinberg, and Takashi Yanagawa for their helpful comments and support during the preparation of this manuscript.

REFERENCES

1. Baan, R. A., Fichtinger-Schepman, A. M. J., Roza, L., and van der Schans, G. P. Molecular dosimetry of genotoxic damage. Arch. Toxicol. (Suppl.) 13: 66–82 (1989).
2. Malkinson, A. M., Nesbitt, M. N., and Skamene, E. Susceptibility to urethane-induced pulmonary adenomas between A/J and *C57BL/6J* mice: use of AXB and BXA recombinant inbred lines indicating a three-locus model. J. Natl. Cancer Inst. 75: 971–1974 (1985).
3. Bora, K. C., Douglas, G. R., and Nestmann, E. R., Eds. Chemical Mutagenesis, Human Population Monitoring, and Genetic Risk Assessment. Elsevier Biomedical Press, Amsterdam, 1982.
4. Ryan, J., Barker, P. E., Nesbitt, M. N., and Ruddle, F. H. KRAS2 as a genetic marker for lung tumor susceptibility in inbred mice. J. Natl. Cancer Inst. 79: 1351–1357 (1987).

5. Caporaso, N. E., Tucker, M. A., Hoover, R. N., Hayes, R. B., Pickle, L. W., Issaq, H. J., Muschik, G. M., Green-Gallo, L., Buivys, D., Aisner, S., Resau, J. H., Trump, B. J., Tollerud, D., Weston, A., and Harris, C. C. Lung cancer and the debrisoquine metabolic phenotype. J. Natl. Cancer Inst. 82: 1264–1272 (1990).
6. Sugimura, H., Caporaso, N. E., Modaii, R. V., Hoover, R. N., Resau, J. H., Trump, B. J., Lonergan, J. A., Krontiris, T. G., Mann, D L., Weston, A., and Harris, C. C. Association of rare alleles of the Harvey *ras* proto-oncogene locus with lung cancer. Cancer Res. 50: 1857–1862 (1990).
7. Tamai, S., Sugimura, H., Caporaso, N. E, Resau, J. H., Trump, B. F Weston, A., and Harris, C. C. Restriction fragment length polymorphism analysis of the L-*myc* gene locus in a case-control study of lung cancer. Int. J. Cancer 46: 411–415 (1990).
8. Fearon, E. R., Feinberg, A. P., Hamilton, S. H., and Vogelstein, B. Loss of genes on the short arm of chromosome 11 in bladder cancer. Nature 318: 377–380 (1985).
9. Cartwright, R. A., Phillip, P. A., Rogers, H. J., and Glashan, R. W. Genetically determined debrisoquine oxidation capacity in bladder cancer. Carcinogenesis 5: 1191–1192 (1984).
10. Bishop, J. M. The molecular genetics of cancer. Science 235: 305–311 (1987).
11. Cavenne, W. K., Scrable, H. J., and James, C. D. Molecular genetics of human cancer predisposition and progression. Mutat. Res. 247: 199–202 (1991).
12. Burmer, G. C., Rabinovitch, R. S., and Loeb, L. A. Analysis of c-Ki-*ras* mutations in human colon carcinoma by cell sorting, polymerase chain reaction, and DNA sequencing. Cancer Res. 49: 2141–2146 (1989).
13. Pohl, H., and Reidy, J. A. Vitamin C intake influences the bleomycin-induced chromosome damage assay: implications for detection of cancer susceptibility and chromosome breakage syndromes. Mutat. Res. 224: 247–252 (1989).
14. Cooper, D. N., and Clayton, J. F. DNA polymorphism and the study of disease associations. Hum. Genet. 78: 299–312 (1988).
15. Khoury, M. J., Flanders, W. D., Greenland, S., and Adams, M. J. On the measurement of susceptibility in epidemiologic studies. Am. J. Epidemiol. 129: 183–190 (1989).
16. Taylor, J. A. Epidemiologic studies of the molecular genetics of cancer. In: Genetic Susceptibility to Environmental Mutagens and Carcinogens (A. D. Bloom, L. Spatz, and N. W. Paul, Eds.), March of Dimes Birth Defects Foundation, White Plains, NY, 1989, pp. 83–93.
17. Johnson, N. L., and Kotz, S. Distributions in Statistics: Discrete Distributions. Houghton-Mifilin, Boston, 1969.
18. van Houwelingen, J. C., and le Cressie, S. Logistic regression, a review. Stat. Neerland. 42: 215–232 (1988).
19. McCullagh, P. Generalized linear models. Eur. J. Operat. Res. 16: 285–292 (1984).
20. SAS Institute Inc. SAS User's Guide: Statistics, Version 5 ed. Cary, NC, 1985.
21. Payne, C. D., Ed. The GLIM System, Release 3.77, Manual. Numerical Algorithms Group, Oxford, 1987.
22. Hommel, G. A stagewise rejective multiple test procedure based on a modified Bonferroni test. Biometrika 75: 383–386 (1988).
23. Kotz, S., Johnson, N. L., and Read, C. B. Wald's *W*-statistics. In: Encyclopedia of Statistical Sciences, Vol. 9 (S. Kotz, N. L. Johnson, and C. B. Read, Eds.), Wiley, New York, 1988, pp. 525–526.
24. Piegorsch, W. W. One-sided significance tests for generalized linear models under dichotomous response. Biometrics 46: 309–316 (1990).
25. Richardson, J. T. E. Variants of chi-square for 2 × 2 contingency tables. Br. J. Math. Stat. Psychol. 43: 309–326 (1990).
26. Haseman, J. K. Exact sample sizes for use with the Fisher-Irwin test for 2 × 2 tables. Biometrics 34: 106–109 (1978).
27. Gart, J. J. The comparison of proportions: a review of significance tests, confidence intervals and adjustments for stratification. Rev. Int. Stat. Inst. 39: 148–169 (1971).
28. Ottman, R. An epidemiologic approach to gene-environment interaction. Genet. Epidemiol. 7: 177–185 (1990).
29. Breslow, N., and Day, N. Statistical Methods in Cancer Research. I. The Analysis of Case-Control Studies. IARC Scientific Publications, International Agency for Reaearch on Cancer, Lyon, France, 1980.
30. Gardner, M. J., and Munford, A. G. The combined effect of two factors on disease in a case-control study. Appl. Stat. 29: 276–281 (1980).
31. Bartlett, M. S. Contingency table interaction. J. R. Stat. Soc. (Suppl.) 2: 248–252 (1935).
32. Mood, A. M., Graybill, F. A., and Boes, D. C. Introduction to the Theory of Statistics. McGraw-Hill, New York, 1974.
33. Taylor, J. A. Epidemiologic evidence of genetic susceptibility to cancer. Birth Defects 26: 113–127 (1990).
34. Kleinbaum, D. G., Kupper, L. L., and Morgenstern, H. Epidemiologic Research: Principles and Quantitative Methods. Lifetime Learning Publications, Belmont, CA, 1982.
35. Anderson, J. A. Separate sample logistic discrimination. Biometrika 59: 19–35 (1972).
36. Prentice, R. L., and Pyke, R. Logistic disease incidence models and case control studies. Biometrika 66: 403–411 (1979).
37. Scott, A. J., and Wild, C. J. Hypothesis testing in case-control studies. Biometrika 76: 806–808 (1989).
38. Caporaso, N., Hayes, R. B., Dosemeci, M., Hoover, R., Ayesh, R., Hetzel, M., and Idle, J. Lung cancer risk, occupational exposure, and the debrisoquine metabolic phenotype. Cancer Res. 49: 3675–3679 (1989).
39. Nebert, D. W. Polymorphism of human *CYP2D* genes involved in drug metabolism: possible relationship to individual cancer risk. Cancer Cells 3: 93–96 (1991).
40. Finkelstein, D. M., and Schoenfeld, D. A. Analysis of multiple tumor data from a rodent carcinogenicity experiment. Biometrics 45: 219–230 (1989).
41. Collings, B. J., Margolin, B. H., and Oehlert, G. W. Analyses for binomial data, with applications to the fiuctuations test for mutagenicity. Biometrics 37: 775–794 (1981).
42. Sobel, M., and Elashoff., R. M. Group testing with a new goal, estimation. Biometrika 62: 181–193 (1975).
43. Chen, C. L., and Swallow, W. H. Using group testing to estimate a proportion, and to test the binomial model. Biometrics 46: 1035–1046 (1990).
44. Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. Discrete Multivariate Analysis: Theory and Practice. MIT Press, Cambridge, MA, 1975.
45. Hansen, E. W. The development of initial and infant behavior in the rhesus monkey. Behavior 27: 107–149 (1966).
46. Cochran, W. G. Estimation of bacterial densities by means of the most probable number. Biometrics 6: 105–116 (1950).
47. McCrady, M. H. The numerical interpretation of fermentation-tube results. J. Infect. Dis. 17: 183–212 (1915).