Evaluating Goodness of Fit of Poisson Regression Models in Cohort Studies

EDWARD L. FROME and MAX D. MORRIS*

Relative and absolute risk models are often used in epidemiologic studies to describe the effect of exposure on age-specific mortality rates. Poisson regression analysis is used to obtain maximum likelihood estimates of unknown parameters and to assess goodness of fit of the models. It is common practice to base both estimation and the evaluation of goodness of fit on certain marginal totals. This approach is potentially misleading for relative risk models and is totally inappropriate for absolute risk models. To illustrate the situation, we present an example using data on smoking and lung cancer. In the example, the goodness-of-fit tests based on the marginal totals indicate that neither the absolute risk nor the relative risk model can be rejected, whereas the age-specific test statistics strongly reject both models. The validity of the marginal test is based on an assumption (i.e., of no exposure by age interaction) that is not satisfied in the example. We describe a more general approach to model evaluation that provides the analyst with summary information for various models of interest and provides guidance on how to identify good descriptive models.

KEY WORDS: Absolute risk models; Epidemiologic follow-up study; Relative risk model.

1. INTRODUCTION

The analysis of censored failure time data is widely used to evaluate the potential adverse health effects of protracted exposure to risk factors. In epidemiologic follow-up studies, large groups of individuals are assembled and followed forward in time. The age at entry, calendar year of entry, and the intensity and duration of exposure to one or more risk factors are typical variables that are determined for each individual. The dependent variable is the time to failure, and this is often taken to be death from a disease of interest (e.g., lung cancer). The fundamental measure of disease occurrences is the hazard or failure rate (i.e., the incidence rate or force of mortality). For the ith individual, the rate $\lambda(t, \mathbf{z}_i(t), \boldsymbol{\beta})$ at age t depends on an m-dimensional vector of known covariates $\mathbf{z}_i(t)$ whose components summarize that individual's exposure history up to age t and the unknown parameter vector $\boldsymbol{\beta}$. The covariate vector may also include demographic and personal characteristics that should be considered as potential confounding factors. In principle, it is possible to use statistical methods that have been developed for the analysis of censored survival data with time-dependent covariates to analyze large cohort studies. A closely related approach that is widely used in practice is to treat all covariates as discrete variables with a small number of levels and partition age into J disjoint intervals. The data are then summarized into a multidimensional table that contains the number of failures (y_{jk}), the "person-years" (c_{jk}), and the value of the covariate vector $\mathbf{z}_{jk}$ for the jth age interval and the kth state. There will be a separate state for each combination of discrete covariate levels in this "grouped data" approach to survival analysis. The general procedure that is used for the person-years calculations when one or more of the covariates is time dependent was described by Clayton (1982) and Breslow, Lubin, Lubin, Marek, and Langholz (1983). The c_{jk} are assumed to be known constants and the y_{jk} are treated as if they were the observed values of independent Poisson variates with means $\mu_{jk} = c_{jk}\,\lambda(\mathbf{z}_{jk}, \boldsymbol{\beta})$; see Holford (1980), Laird and Olivier (1981), Breslow et al. (1983), and Whittemore (1985). The rate function $\lambda(\mathbf{z}, \boldsymbol{\beta})$ is used to model the effect of the covariates on the age-specific mortality rates and plays the role of the regession function in standard regression analysis. Poisson regression analysis is used for estimation, hypothesis testing, and regression diagnostics (Frome 1983; Frome and Checkoway 1985). In many situations, it is convenient to assume that $\lambda(\mathbf{z}, \boldsymbol{\beta})$ can be expressed as the product of a baseline rate λ_j^0 and a relative risk function; that is,

$$\lambda(\mathbf{z}_{jk}, \boldsymbol{\beta}) = \lambda_j^0\, \rho(\mathbf{z}_{jk}, \boldsymbol{\beta}),$$

where the relative risk function ρ is nonnegative and satisfies $\rho(\mathbf{0}, \boldsymbol{\beta}) = 1$ and $\mathbf{z}$ is defined so that $\mathbf{z} = \mathbf{0}$ corresponds to the baseline population. Two special cases of general interest are $\rho(\mathbf{z}, \boldsymbol{\beta}) = 1 + \mathbf{z}\boldsymbol{\beta}$ (additive relative risk) and $\rho(\mathbf{z}, \boldsymbol{\beta}) = \exp(\mathbf{z}\boldsymbol{\beta})$ (multiplicative relative risk), where $\mathbf{z}\boldsymbol{\beta}$ denotes inner product.

In the example that follows, we will consider only the situation described by Crump and Allen (1985), where there is a single covariate with a linear dose–response function that is expressed in an absolute or relative form with respect to λ_j^0, the baseline rates (usually obtained from an external control group, e.g., national vital statistics). In this situation, the mean of the Poisson variate y_{jk} is then given by

$$\mu_{jk} = c_{jk}[\lambda_j^0 + \beta_a z_{jk}] = \mu_{jk}^0 + \beta_a c_{jk} z_{jk} \qquad (1)$$

for the absolute risk model and

$$\mu_{jk} = c_{jk}\, \lambda_j^0(1 + \beta_r z_{jk}) = \mu_{jk}^0(1 + \beta_r z_{jk}) \qquad (2)$$

for the relative risk model.

2. EXAMPLE

The data in Table 1 provide an interesting example of a typical situation in which the absolute risk (1) and relative

*Edward L. Frome and Max D. Morris are both statisticians in the Mathematical Sciences Section, Engineering Physics and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-8083. This work was supported by Office of Health and Enviornmental Research (U.S. Department of Energy) Contracts DE-AC05-84OR21400 with Martin Marietta Energy Systems and DE-AC05-76OR00033 with Oak Ridge Associated Universities. The authors thank T. J. Mitchell, an associate editor, and two referees for helpful comments.

 © 1989 American Statistical Association

Table 1. Lung Cancer Mortality Data From Study of British Physicians

Age group		Cigarettes per day (median value) 0 (0)	1–9 (5.2)	10–14 (11.2)	15–19 (15.9)	20–24 (20.4)	25–34 (27.4)	35+ (40.8)
35–40	y	1	0	0	0	0	0	0
	μ^0	.091	.028	.032	.038	.050	.027	.006
	c	10,366	3,121	3,577	4,317	5,683	3,042	670
40–45	y	0	0	1	0	1	1	0
	μ^0	.218	.078	.088	.113	.171	.108	.031
	c	8,162	2,937	3,286	4,214	6,385	4,050	1,166
45–50	y	0	0	1	0	1	4	0
	μ^0	.332	.127	.142	.177	.305	.239	.082
	c	5,969	2,288	2,546	3,185	5,483	4,290	1,482
50–55	y	0	0	2	4	6	9	4
	μ^0	.407	.182	.201	.232	.424	.386	.143
	c	4,496	2,015	2,219	2,560	4,687	4,268	1,580
55–60	y	0	1	0	0	5	9	6
	μ^0	.453	.213	.236	.244	.470	.455	.172
	c	3,512	1,648	1,826	1,893	3,646	3,529	1,336
60–65	y	0	2	1	2	12	11	10
	μ^0	.385	.229	.243	.233	.422	.424	.162
	c	2,201	1,310	1,386	1,334	2,411	2,424	924
65–70	y	0	0	2	2	9	10	7
	μ^0	.201	.197	.210	.180	.332	.299	.118
	c	1,421	927	988	849	1,567	1,409	556
70–75	y	0	3	4	2	7	5	4
	μ^0	.237	.150	.144	.099	.181	.140	.054
	c	1,121	710	684	470	857	663	255
75–80	y	2	0	3	5	7	3	1
	μ^0	.405	.297	.220	.137	.204	.139	.051
	c	826	606	449	280	416	284	104
	$y_{\cdot k}$	3	6	14	15	48	52	32
	$\mu_{\cdot k}$	2.829	1.501	1.513	1.453	2.559	2.217	.819
	$c_{\cdot k}$	38,074	15,562	16,916	19,102	31,135	23,959	8,073

NOTE: y represents lung cancer deaths, μ^0 represents expected deaths based on an external control group, and c represents person-years at risk.

risk (2) regression functions are of interest. These data are part of a study of the health effects of cigarette smoking in British physicians (Doll and Hill 1966) and have been used by Frome (1983) to illustrate how Poisson rate analysis can be used to model nonlinear dose–response relations. Table 1 has been restricted to include only those regular smokers who started smoking at about age 20 and lifelong nonsmokers (column 1).

Smokers were categorized according to their *exposure rate*, expressed as daily cigarette consumption (cigarettes per day). The average cigarette consumption, say d_k, in each of the "dose" groups is given at the top of each column. Doll (1971) suggested that the lung cancer data for nonsmokers reported by Hammond (1966) constitutes an appropriate external control group. We have used these data (Hammond 1966, table 22 and app. table I) to obtain smoothed estimates (based on a cubic spline for the log-rates) of the age-specific lung cancer death rates that were used to compute the μ_{jk}^0 values shown in Table 1.

Table 2 shows the values of the deviance for the absolute risk model (1) and relative risk model (2) for the age-specific data in Table 1. The deviance is a likelihood ratio statistic used as a measure of unexplained variation similar to the residual sum of squares and is related to the log-likelihood function by $D(\hat{\boldsymbol{\beta}}) = -2[L(\hat{\boldsymbol{\beta}}) - L(\mathbf{y})]$, where $L(\hat{\boldsymbol{\beta}})$ denotes the log-likelihood function evaluated at the maximum likelihood (ML) estimate $\hat{\boldsymbol{\beta}}$ and $L(\mathbf{y})$ is the log-likelihood for the saturated model. If the assumed regression function is appropriate, the deviance will be distributed approximately as a chi-square with 62 df. The results for the age-specific data in Table 2 indicate that both models suffer significant lack of fit, but suggest that the relative risk model provides a better descriptive summary of these data than the absolute risk model.

A practice often followed in epidemiologic cohort studies is to estimate the regression parameters using certain marginal totals (see, e.g., Crump and Allen 1985; Howe, Nair, Newcombe, Miller, and Abbatt 1986; Land, Boice, Shore, Norman, and Tokunaga 1980). This practice will sometimes lead to correct estimates (i.e., the ML estimates that would be obtained from the stratified data if age is not a confounder or an effect modifier, but may not be appropriate if the goodness of fit of the model is being evaluated. For each of the K exposure groups the total deaths $y_{\cdot k} = \Sigma_j y_{jk}$ is a Poisson variate with mean $\mu_{\cdot k}^0 + \beta_a c_{\cdot k} d_k$ for the absolute

Table 2. Deviance Values for Table 1 Data

Model	Age-specific data	Marginal totals
Absolute risk	289.5	5.96
Relative risk	95.9	2.83
Degrees of freedom	62	6

Table 3. Hypotheses Tested in Table 2

Model	(a) Using age-specific data	(b) Using marginal totals
Absolute risk	$H_0 : \mu^0_{jk} + \beta_a c_{jk} d_k$ $H_a : \mu_{jk}$	$H_0 : \mu^0_{\cdot k} + \beta_a c_{\cdot k} d_k$ $H_a : \mu_{\cdot k}$
Relative risk	$H_0 : \mu^0_{jk}(1 + \beta_r d_k)$ $H_a : \mu_{jk}$	$H_0 : \mu^0_{\cdot k}(1 + \beta_r d_k)$ $H_a : \mu_{\cdot k}$

risk model and mean $\mu^0_{\cdot k}(1 + \beta_r d_k)$ for the relative risk model. Table 2 shows the deviance values for both of these models obtained using the marginal totals (see bottom of Table 1). Since both of these deviance values are small (compared with a chi-square with 6 df), *neither* the absolute nor the relative risk model appears to be inappropriate.

In summary, the results in Table 2 that were obtained using marginal totals would lead us to believe that neither the absolute nor relative risk models should be rejected. In contrast, when the age-specific rates were used we concluded that both models are inadequate. The apparent discrepancy can be better understood in terms of the specific hypotheses being compared. These hypotheses are listed in Table 3. The null and alternative hypotheses of Table 3b are implied by the corresponding hypotheses of Table 3a. If the null hypothesis is true in the formulation involving the entire table, then the corresponding null hypotheses for the marginal test is also true. (This is the basis for the "validity" of the marginal tests.) A null hypothesis in Table 3b can be true, however, while the corresponding null hypothesis in Table 3a is false. For example, the relative risk model stipulates that a linear form in d_k describes the proportional effect of exposure. This may be true overall (where age is not considered), but the same linear form may not be adequate when applied to each age stratum individually. Hence the value of the deviance for the test in Table 3a could indicate significant lack of fit, whereas the value of the deviance (based on marginal totals) for the corresponding test in Table 3b does not. On the other hand, suppose that the inadequacy is due to a quadratic term in the relative risk function that is common to each row; that is, $\mu_{jk} = \mu^0_{jk}(1 + \beta_r d_k + \gamma_r d_k^2)$. Then the test of Table 3b should be more powerful than the one in Table 3a, since it is based upon fewer degrees of freedom. Thus the test in Table 3a might indicate no significant lack of fit, whereas the corresponding test in Table 3b might indicate the contrary.

A less formal, but perhaps more useful, approach to model selection can be based upon an extended table that describes various models of interest and gives the values of the deviance and degrees of freedom for each model. This approach is useful when several possibly nonnested models are being considered and will help to explain the apparent inconsistency between the age-specific and marginal analyses that was illustrated in Table 2. The goodness-of-fit tests in Table 2 are, in fact, likelihood ratio tests with different alternative hypotheses. To see this, consider Table 4, which shows the deviance values for several relative risk models of interest. The first line gives the value of the deviance for the minimal model in which the excess relative risk is constant. Line 2 is the relative risk model (2) with $z_{jk} = d_k$, and on line 3 we assume that $z_{jk} = t_j d_k$, where t_j is duration of smoking (age minus 20), so z_{jk} is the cumulative exposure. Line 4 corresponds to the model

$$\mu_{jk} = \mu^0_{jk}[1 + \beta t_j^{\phi} d_k], \tag{3}$$

where the excess risk is proportional to the baseline rates, the exposure rate, and the duration of exposure to the power ϕ. Note that the models on lines 2 and 3 are special cases of (3) obtained by setting $\phi = 0$ or 1. The fifth and sixth lines are factor models for column effects (exposure rate) and row effects (age), respectively. The deviance for a two-factor model, where the excess risk is represented as the product of row and column effects, is given on line 7. Lines 3, 4, and 7 all represent exposure by age interaction models.

The deviance values and degrees of freedom in Table 4 can be used to compute the value of various likelihood ratio test statistics of interest (see Table 5). In Tables 4 and 5, the models are listed in increasing order of complexity. In Table 5, we have computed the significance level for each goodness-of-fit test for each model in Table 4. The null hypothesis is given at the top of each column, with the alternative hypothesis listed on the row. For example, the lack-of-fit chi-squared value for the linear exposure rate $\mu^0_{jk}(1 + \beta d_k)$ model with the unconstrained column effects model $\mu^0_{jk}(1 + \delta_k)$ as the alternative is 2.83 with 6 df (see lines 2 and 5 in Table 4). This is the same test statistic that was obtained using the marginal totals (see Table 2), and the significance level (.83) is given in column 1 of row 2 in Table 5. This is the only alternative model that would not cause us to strongly reject the linear exposure rate model. Note, however, that when the unconstrained column effects model is tested against the alternative $\mu^0_{jk}(1 + \alpha_j \delta_k)$, this model [i.e., $H_0 : \alpha_j = 1$ ($j = 1, \ldots, 9$)] is strongly rejected (see row 3, column 4 of Table 5).

The simplest model in Table 4 that cannot be rejected for at least one alternative model (see column 3 in Table 5) is $\mu^0_{jk}(1 + \beta t_j^{\phi} d_k)$, and consequently we might call this the "best" model for these data. Note that when this model is used to define the alternative hypothesis both the exposure rate model ($\phi = 0$) and the cumulative exposure model ($\phi = 1$) are obtained as special cases and are strongly rejected (see row 1 of Table 5).

Another less formal approach that can be used to identify

Table 4. Deviance Values for Relative Risk Models—$\mu^0_{jk}[1 + \eta_{jk}]$

Excess risk—η_{jk}	df	Deviance	AIC*
η	62	209.56	211.56
βd_k	62	95.90	97.90
$\beta t_j d_k$	62	70.73	72.73
$\beta t_j^{\phi} d_k$	61	63.83	67.83
δ_k exposure rate	56	93.07	107.07
α_j age	54	170.17	188.17
$\alpha_j \delta_k$	49	51.01	79.01

*AIC = deviance + 2(number of parameters); see text.

Table 5. Significance Levels for Likelihood Ratio Tests for Relative Risk Model in Table 4

Alternative hypothesis	Null hypothesis, βd_k	$\beta t_j d_k$	$\beta t_j^{\phi} d_k$	δ_k	$\alpha_j \delta_k$
$\beta t_j^{\phi} d_k$	<.0001	.0086	—	—	—
δ_k	.8299	—	—	—	—
$\alpha_j \delta_k$	.0001	.1024	.3923	<.0001	—
μ_{jk}	.0037	.2092	.3773	.0014	.3945

"good models" is based on the Akaike information criterion (AIC),

$$\text{AIC} = D(\hat{\boldsymbol{\beta}}) + 2m,$$

where $D(\hat{\boldsymbol{\beta}})$ is the deviance and m is the number of parameters for the specified model. This expression combines a measure of the discrepancy between the fitted values $\hat{\mu}_{jk}$ and the data (as measured by the deviance) and the simplicity of the model as reflected in the number of parameters (see Atkinson 1981; McCullagh and Nelder 1983). Results given by Efron (1986) show that using AIC to compare competing models with different numbers of parameters amounts to selecting the model with the minimum estimate of the mean squared error of prediction. The value of AIC for each model in Table 4 is shown in the last column. Note that the two models in Table 4 that involve duration of exposure and intensity of exposure have the smallest AIC values, and this, together with the results in Table 5, suggests that Equation (3) is the best descriptive model for the data in Table 1.

In our example, we were unable to identify any absolute risk models that provide a good fit and have, therefore, limited our presentation in Table 4 to relative risk models. In principle, Table 4 could be expanded to include absolute risk models, and Muirhead and Darby (1987) have provided an approach to defining more general regression functions that include absolute risk and relative risk models as special cases. It should be noted, however, that for absolute risk models, the ML estimates obtained from the marginal tools will not agree with those obtained for the stratified data. The fact that relative risk models fitted to summary data yield correct estimates, whereas additive risk models do not, follows immediately upon consideration of the sufficient statistics for the full data case. There are also additional regression functions (that do not use external rates) that have been used to describe the age-specific lung cancer death rates for the data in Table 1 and related data sets (see, e.g., Breslow 1985; Frome 1983; Whittemore and Altshuler 1976).

3. CONCLUSION

The main purpose of this article has been to focus attention on the methods that are used to evaluate the goodness of fit of dose–response models in epidemiologic follow-up studies. This is an important issue, since these models provide a basis for evaluating the potential adverse health effects of occupational and environmental exposure to hazardous substances. In some situations the resulting models are used to extrapolate risk estimates or to determine "safe" levels of exposure. The example that we presented clearly shows that conflicting results can occur when marginal totals are used (instead of age-specific mortality rates) to evaluate goodness of fit. To illustrate why this happens, we presented a more detailed analysis, which shows that the choice of the alternative hypothesis is important. When specific models can be represented as a special case of a more general model, then the well-known Neyman–Pearson theory leads to likelihood ratio test statistics. This approach cannot be used for nonnested models, and we suggest that the AIC can be used to identify good models in this situation.

[*Received March 1988. Revised January 1989.*]

REFERENCES

Atkinson, A. C. (1981), "Likelihood Ratios, Posterior Odds and Information Criteria," *Journal of Econometrics*, 15, 15–20.

Breslow, N. E. (1985), "Cohort Analysis in Epidemiology," in *A Celebration of Statistics*, eds. A. C. Atkinson and D. E. Fienberg, New York: Springer-Verlag, pp. 109–143.

Breslow, N. E., Lubin, J. H., Marek, P., and Langholz, B. (1983), "Multiplicative Models and Cohort Analysis," *Journal of the American Statistical Association*, 78, 1–12.

Clayton, D. G. (1982), "The Analyses of Prospective Studies of Disease Etiology," *Communications in Statistics, Part A*, 11, 2129–2155.

Crump, K. S., and Allen, B. C. (1985), "Methods for Quantitative Risk Assessment Using Occupational Studies," *The American Statistician*, 39, 442–450.

Doll, R. (1971), "The Age Distribution of Cancer: Implications for Models of Carcinogenesis," *Journal of the Royal Statistical Society*, Ser. A, 135, 133–166.

Doll, R., and Hill, A. B. (1966), "Mortality of British Doctors in Relation to Smoking: Observations on Coronary Thrombosis," in *Epidemiologic Study of Cancer and Other Chronic Diseases* (National Cancer Institute Monograph 19), ed. W. Haenszel, Washington, DC: U.S. Government Printing Office, pp. 205–258.

Efron, B. (1986), "How Biased Is the Apparent Error Rate of a Prediction Rule?" *Journal of the American Statistical Association*, 81, 461–470.

Frome, E. L. (1983), "The Analysis of Rates Using Poisson Regression Models," *Biometrics*, 39, 665–675.

Frome, E. L., and Checkoway, H. (1985), "Use of Poisson Regression Models in Estimating Incidence Rates and Ratios," *American Journal of Epidemiology*, 121, 309–323.

Hammond, E. E. (1966), "Smoking in Relation to Death Rates of One Million Men and Women," in *Epidemiologic Study of Cancer and Other Chronic Diseases* (National Cancer Institute Monograph 19), ed. W. Haenszel, Washington, DC: U.S. Government Printing Office, pp. 127–204.

Holford, T. R. (1980), "The Analysis of Rates and Survivorship Using Log-Linear Models," *Biometrics*, 36, 229–305.

Howe, G. R., Nair, R. C., Newcombe, H. B., Miller, A. B., and Abbatt, J. D. (1986), "Lung Cancer Mortality (1950–80) in Relation to Radon Daughter Exposure in a Cohort of Workers at the Eldorado Beaverlodge Uranium Mine," *Journal of the National Cancer Institute*, 77, 357–362.

Laird, N., and Olivier, D. (1981), "Covariance Analysis of Censored Survival Data Using Log-Linear Analysis Techniques," *Journal of the American Statistical Association*, 76, 231–240.

Land, C. E., Boice, J. D., Shore, R. E., Norman, J. E., and Tokunaga, M. (1980), "Breast Cancer Risk From Low-Dose Exposure to Ionizing Radiation: Results of Parallel of Three Exposed Populations of Women," *Journal of the National Cancer Institute*, 65, 353–376.

McCullagh, P., and Nelder, J. A. (1983), *Generalized Linear Models*, London: Chapman & Hall.

Muirhead, C. R., and Darby, S. C. (1987), "Modelling the Relative and Absolute Risk of Radiation Induced Cancers," *Journal of the Royal Statistical Society*, Ser. A, 150, Part 2, 83–118.

Whittemore, A. (1985), "Analyzing Cohort Mortality Data," *The American Statistician*, 39, 437–441.

Whittemore, A., and Altshuler, B. (1976), "Lung Cancer Incidence in Cigarette Smokers: Further Analysis of Doll and Hill's Data for British Physicians," *Biometrics*, 32, 805–816.