Model-based Estimation of Population Attributable Risk under Cross-sectional Sampling

Srabashi Basu[1] and J. Richard Landis[2]

The covariate-adjusted population attributable risk (PAR) measures the proportionate reduction in disease prevalence in the target population when the putative risk factor is removed. after adjusting for covariate effects. This paper extends the model-based approach developed for retrospective and cohort studies to the cross-sectional sampling design. An appropriate logit linear model is utilized to estimate the covariate-adjusted attributable risk. The asymptotic variance of this complex ratio estimate is obtained using Taylor series expansions which incorporate the sampling variation of the estimated model parameters and the appropriate estimates of risk factor prevalence. These methods are illustrated with cardiovascular disease risk factor data from the second National Health and Nutrition Examination Survey (NHANES II). *Am J Epidemiol* 1995;142:1338–43.

logit models; population attributable risk; statistics

Quantification of the impact of exposure to a risk factor on a particular disease in a target population is a primary public health concern. The population attributable risk (PAR) ratio (1) was introduced to quantify the impact of a binary risk factor on a binary disease outcome. Walter (2–5) derived variance estimates of the PAR under retrospective, prospective. and cross-sectional sampling designs for this simplest situation. More often than not, however, the risk exposure is multifactorial. In addition to the primary risk factor, a number of confounding variables may be present in the population that interact with the risk and/or with the disease response. To adjust for a covariate C present at K levels in the population, 0,. . ., $K - 1$, a covariate-adjusted population attributable risk may be defined as

$$\lambda_A = 1 - \sum_{k=0}^{K-1} [\Pr(C_k)\Pr(D_1 | E_0 C_k)]/[\Pr(D_1)], \qquad (1)$$

where the subscript 0 (1) denotes absence (presence) for D corresponding to disease status and E to exposure status (4–9).

A unified formulation of PAR in the general multivariable setting with more than one risk factor has been put forward for retrospective and prospective studies (10–18). Except for the simplest case of a binary risk factor and a binary disease, no variance formulation for PAR under cross-sectional sampling has been developed. Unlike a prospective study where the exposure prevalences are assumed known, the estimate of PAR as well as the estimate of its variance, under cross-sectional sampling must incorporate the estimated exposure prevalences. In this paper, we propose an extension of existing methodology (10, 11, 17, 18) to estimate PAR for polychotomous risk factors and covariates under cross-sectional sampling. Its asymptotic variance has been formulated by accounting for the additional source of variability introduced by the estimated exposure prevalences.

In the next section, an example is introduced using data from the second National Health and Nutrition Examination Survey (NHANES II). In the example, the disease of interest is hypertension among young adult women. We then describe the methodology and the estimation procedure in detail. Under Results, we present the theoretical results in the context of the NHANES II data. A discussion follows in the final section.

Received for publication May 25, 1994, and in final form March 27, 1995.

Abbreviations: BMI, body mass index; DBP, diastolic blood pressure; NHANES II, second National Health and Nutrition Examination Survey; PAR, population attributable risk; SE, standard error.

[1] Indian Statistical Institute, Calcutta, West Bengal, India.

[2] Center for Biostatistics and Epidemiology, Penn State University College of Medicine, Hershey, PA.

Reprint requests to Dr. J. Richard Landis, Center for Biostatistics and Epidemiology, Penn State University College of Medicine, 500 University Drive, Room C20705, Hershey, PA 17033.

EXAMPLE

Consider the data presented in table 1, which were obtained from NHANES II conducted from 1976 to 1980 (19). These data were selected from a larger research project investigating secular trends in cardiovascular disease risk factors over the 20-year period 1960–1980 in the United States (20). These unweighted frequency data summarize the distribution of diastolic blood pressure (DBP) among young adult women, aged 18–24 years, by ordinal levels of body mass index and race. Although these data were obtained from a weighted, complex survey design, the assumption of simple random sampling is used throughout this paper for simplification. Arbitrarily, DBP exceeding 82.6 mmHg (determined from the weighted 90th percentile of the distribution) is regarded as disease being present for these cross-sectional survey data. An important public health consideration is the differential impact of body mass index (BMI), expressed as weight (kg)/height (m)^2, and race on elevated blood pressure.

In the next section, we develop a methodology to assess the impact of BMI and race on elevated DBP measurements, taking into account the fact that the underlying sampling design is cross-sectional. From a primary prevention perspective, we consider BMI as the primary risk factor and race as a potential covariate influencing hypertension.

TABLE 1. Frequency distribution (row proportions) of diastolic blood pressure (DBP) at two levels by multiple levels of body mass index (BMI) (kg/m^2) and race: women, aged 18–24 years, NHANES II*

Race	BMI	DBP >90th percentile		Total	Estimated factor prevalences
		Yes (D_1)	No (D_0)		
Black (C_1)	≥27 (E_3)	10 (0.3846)	16 (0.6154)	26 (1.00)	0.0269
	[25,27) (E_2)	1 (0.0769)	12 (0.9231)	13 (1.00)	0.0135
	[23,25) (E_1)	3 (0.1429)	18 (0.8571)	21 (1.00)	0.0217
	<23 (E_0)	10 (0.1370)	63 (0.8630)	73 (1.00)	0.0756
	Subtotal	24	109	133	0.1377
White (C_0)	≥27 (E_3)	29 (0.3118)	64 (0.6882)	93 (1.00)	0.0963
	[25,27) (E_2)	7 (0.1094)	57 (0.8906)	64 (1.00)	0.0662
	[23,25) (E_1)	8 (0.0734)	101 (0.9266)	109 (1.00)	0.1128
	<23 (E_0)	40 (0.0705)	527 (0.9294)	567 (1.00)	0.5869
	Subtotal	84	749	833	0.8623
Total		108	858	966	1.0000

* NHANES II, second National Health and Nutrition Examination Survey.

METHODS

Although logistic regression modeling is general enough to be utilized as a framework to produce estimates of covariate-adjusted population attributable risk across a wide range of situations, we chose to formulate the methods within the simpler context of a primary risk factor reported on I levels and a single binary covariate. Let $L^c(i)$ be the simple logit at exposure level i and covariate level c, and γ be the coefficient for the covariate status. Let

$$\text{Model I: } L^c(i) = \alpha + \beta_i + c\gamma, \quad i = 0, \ldots, I-1, c = 0, 1. \tag{2}$$

where β_i is the effect of the ith level of the primary risk factor reported on a nominal scale. Incorporating the effects of the primary risk factor on an ordinal scale, let

$$\text{Model II: } L^c(i) = \alpha + i\beta + c\gamma, \quad i = 0, \ldots, I-1, c = 0, 1. \tag{3}$$

where β is the trend coefficient. To identify the model parameters uniquely, assume that $\beta_0 = 0$ in Model I. Note
that the only difference between the two models is the replacement of β_i in Model I with $i\beta$ in Model II.
It is useful to give a parametric formulation for the conditional probability of disease for an individual
belonging to the ith level of exposure (E_i) and cth level of covariate (C_c) under these models as

$$\Pr(D_1|E_iC_c) = [\exp(\alpha + \beta_i + c\gamma)]/[1 + \exp(\alpha + \beta_i + c\gamma)]. \tag{4}$$

using Model I for purposes of illustration. Throughout subsequent developments, these formulations will be developed in terms of Model I, but can be applied readily to Model II.

Following equation 1, and replacing the conditional disease probability by its parametric formulation given in equation 4, the covariate-adjusted population attributable risk for multiple risk levels is

$$\lambda_A = 1 - [\sum_{c=0}^{1} \Delta_0 \Pr(C_c)]/[\sum_{c=0}^{1} \sum_{i=0}^{I-1} \Delta_i \Pr(C_c\, E_i)], \tag{5}$$

where $\Delta_i = [\exp(\alpha + \beta_i + c\gamma)]/[1 + \exp(\alpha + \beta_i + c\gamma)]$.

Next we describe the procedure to derive the maximum likelihood estimator for the covariate-adjusted PAR together with its variance under a cross-sectional sampling design. Maximum likelihood estimators of the parameters for models I and II are obtained readily from PROC LOGIST in SAS (SAS Institute, Cary, North Carolina). These estimates and their asymptotic covariance matrix are the same as given in Greenland and Drescher (18). However, under a prospective sampling design, the exposure prevalences are assumed to be known quantities, and are treated as constants for the derivations of the asymptotic variance of the estimated PAR (18). Under cross-sectional sampling, the exposure prevalence must be estimated and the covariance between model parameters and the exposure prevalence is derived by applying the implicit function theorem (21).

The maximum likelihood estimators of the factor prevalence rates are the observed proportions in each of the covariate-exposure categories. An estimate of the covariate-adjusted PAR under the cross-sectional sampling is given by equation 5 with all parameters and factor prevalence rates replaced by their maximum likelihood estimators, i.e.,

$$\hat{\lambda}_A = 1 - [\sum_{c=0}^{1} \hat{\Delta}_0 \hat{\nu}_{c+}]/[\sum_{c=0}^{1} \sum_{i=0}^{I-1} \hat{\Delta}_i \hat{\nu}_{ci}], \tag{6}$$

where $\hat{\Delta}_i = [\exp(\hat{\alpha} + \hat{\beta}_i + c\hat{\gamma})]/[1 + \exp(\hat{\alpha} + \hat{\beta}_i + c\hat{\gamma})]$, $\hat{\nu}_{ci} = \hat{\Pr}(C_c E_i)$, and $\hat{\nu}_{c+} = \hat{\Pr}(C_c)$ $i = 1, \ldots, I - 1$, $c = 0, 1$.

Note that $\hat{\lambda}_A$ is a function of two dependent sets of random vectors, namely, the estimated parameter vector, $\mathbf{b} = (\hat{\alpha}, \hat{\beta}_1, \ldots, \hat{\beta}_{I-1}, \hat{\gamma})$, (or $\mathbf{b} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma})$ as the case may be) and the vector of cell proportions, $\mathbf{p}$. The covariance matrix between $\mathbf{b}$ and $\mathbf{p}$ can be expressed as

$$\mathbf{C} = \hat{\mathbf{V}}_{\mathbf{b}} \mathbf{H} \hat{\mathbf{\Sigma}}, \tag{7}$$

where $\hat{\mathbf{V}}_{\mathbf{b}}$ and $\hat{\mathbf{\Sigma}}$ are estimated dispersion matrices of $\mathbf{b}$ and $\mathbf{p}$, respectively, and $\mathbf{H}$ is the Hessian matrix (21). $\hat{\mathbf{V}}_{\mathbf{b}}$ and $\hat{\mathbf{\Sigma}}$ are obtained easily; $\hat{\mathbf{\Sigma}}$ is the observed variance-covariance matrix of a single multinomial random vector, and $\hat{\mathbf{V}}_{\mathbf{b}}$ is the inverse of the observed information matrix. The (l,k)th element of $\mathbf{H}$ is given by $(\partial L/\partial b_l \partial p_k)$, where L is the appropriate log likelihood, $\mathbf{b}_l$ is the lth element of the vector $\mathbf{b}$ and $\mathbf{p}_k$ is the kth element of the vector $\mathbf{p}$.

Let the Jacobian of $\ln(1 - \hat{\lambda}_A)$ with respect to the elements of $\mathbf{b}$ be denoted by $\mathbf{J}^{\mathbf{A}}$ with its lth element given by $(\partial[\ln(1 - \hat{\lambda}_A)]/\partial \mathbf{b}_l)$, and the Jacobian with respect to the estimated cell frequencies be denoted by $\mathbf{B}^{\mathbf{A}}$ with its kth element given by $(\partial[\ln(1 - \hat{\lambda}_A)]/\partial \mathbf{p}_k)$. Therefore

$$\widehat{\mathrm{Var}}_c(\hat{\lambda}_A) = (1 - \hat{\lambda}_A)^2[\mathbf{J}^{\mathbf{A}} \hat{\mathbf{V}}_b \mathbf{J}^{\mathbf{A}'} + \mathbf{B}^{\mathbf{A}} \hat{\mathbf{\Sigma}} \mathbf{B}^{\mathbf{A}'} + 2\mathbf{B}^{\mathbf{A}} \mathbf{C}' \mathbf{J}^{\mathbf{A}'}], \tag{8}$$

where the subscript c designates cross-sectional sampling.

RESULTS

Let us now apply the theoretical results derived above in the context of the data presented in table 1. The covariate-adjusted PAR may be interpreted as that proportion of disease in the target population which theoretically could be eliminated if all exposed individuals revert back to the baseline level of no exposure, after adjusting for the covariate effect. In the absence of any covariate, λ_A measures the reduction of disease in the target population attributable to the primary risk factor only. As a special case of λ_A, we define another quantity measuring the overall reduction in the disease when both the primary risk factor and the covariate are absent in the population. Clearly, if there are I levels of exposure and two levels of confounding, one can form a combined variable with

TABLE 2. **Alternate models for the analysis of NHANES II* data**

Logit model	Covariate in model	BMI* scale	Lack of fit significance level	−2 log likelihood	Model parameters†
No	None	Nominal	–	–	–
Yes	None	Nominal	1.00	628.22	$\hat{\beta}_1 = 0.09$, $\hat{\beta}_2 = 0.31$, $\hat{\beta}_3 = 1.75$
		Ordinal	0.05	635.19	$\hat{\beta} = 0.55$
	Race	Nominal	0.73	624.66	$\hat{\beta}_1 = 0.06$, $\hat{\beta}_2 = 0.28$, $\hat{\beta}_3 = 1.70$ $\hat{\gamma} = 0.52$
		Ordinal	0.22	631.67	$\hat{\beta} = 0.53$ $\hat{\gamma} = 0.51$

* BMI, body mass index; NHANES II, second National Health and Nutrition Examination Survey.
† $\hat{\beta}_i$ is the BMI effect for the ith level relative to the baseline (BMI <23). $\hat{\beta}$ is the BMI trend parameter under equally spaced scoring. $\hat{\gamma}$ is the race effect.

$2I$ levels and λ_A can be used directly to estimate the overall PAR. However, to distinguish between these two quantities for the example considered, let us denote the overall PAR by λ_O.

Two separate modeling strategies are considered in the Methods section. Although the actual sampling design for NHANES II was cross-sectional, the methods for prospective (18) as well as cross-sectional sampling designs are applied to these data to facilitate comparisons. Recall that the form of PAR is equivalent under both designs. However, under the former, the exposure prevalences are taken to be known constants, rendering $\hat{\lambda}_A$ to be a function of **b** only. In other words,

$$\widehat{\mathrm{Var}}_p(\hat{\lambda}_A) = (1 - \hat{\lambda}_A)^2[\mathbf{J}^{\mathbf{A}}\hat{\mathbf{V}}_b\mathbf{J}^{\mathbf{A}'}],$$

where the subscript $_p$ denotes the prospective sampling scheme. In effect, under prospective sampling, the exposure prevalences are taken to be equivalent to their estimated values under cross-sectional sampling to facilitate comparison, but are treated as known constants.

A summary of these results is presented in tables 2 and 3. Table 2 provides essential information for the various models considered.

Recall that BMI is assumed to be the primary risk factor and race the covariate. Using the results in the first part of table 3, a comparison can be made between the model-based and the non-model-based estimates of the proportion of individuals in the target population belonging to the category of DBP ≥90th percentile attributable to BMI ≥23. In order to obtain the non-model-based estimate of PAR, defined only for a 2 × 2 table, the data in table 1 were collapsed over the three highest exposure levels and over race as well. For that situation, labeled "no model," the PAR is estimated to be 30.1 percent, and the standard errors under the prospective and the cross-sectional sampling designs are calculated to be 6.78 and 6.84 percent, respectively (2).

TABLE 3. **Population attributable risk (PAR) analysis of NHANES II* data under alternate models: prospective and cross-sectional designs*, †**

Logit model	Covariate in model	BMI* scale	$\hat{\lambda}_A$	$\mathrm{SE}_p(\hat{\lambda}_A)$	$\mathrm{SE}_c(\hat{\lambda}_A)$	$\hat{\lambda}_O$	$\mathrm{SE}_p(\hat{\lambda}_O)$	$\mathrm{SE}_c(\hat{\lambda}_O)$
No	None	Nominal	0.301	0.0678	0.0684	–	–	–
Yes	None	Nominal	0.301	0.0678	0.0684	–	–	–
		Ordinal	0.368	0.0582	0.0601	–	–	–
	Race	Nominal	0.292	0.0673	0.0692	0.346	0.0680	0.0695
		Ordinal	0.361	0.0589	0.0610	0.409	0.0559	0.0605

* BMI, body mass index; NHANES II, second National Health and Nutrition Examination Survey; SE, standard error.
† $\hat{\lambda}_A$ is the covariate-adjusted population attributable risk parameter. $\hat{\lambda}_O$ is the overall population attributable risk parameter incorporating the combined effects of race and BMI.

In the second panel of table 3, the impact of BMI on DBP is quantified first without considering the effects of the covariate. The nominal scale logit model in this case is

$$\text{Model I: } L(i) = \alpha + \beta_i, \quad i = 0,\ldots, I - 1, \quad (9)$$

and the ordinal scale model is

$$\text{Model II: } L(i) = \alpha + i\beta, \quad i = 0,\ldots, I - 1. \quad (10)$$

In the absence of a covariate in the model, the overall PAR λ_O and the covariate-adjusted PAR λ_A are identical. Under models in equations 9 and 10, the parametric formulation of the PAR follows easily from equation 5 with $c = 0$, yielding PAR estimates to be 30.1 percent and 36.8 percent, respectively. In the last panel of table 3, the results of the covariate-adjusted analyses described in the previous section are summarized. Note that white race is taken to be the reference level for the covariate.

Several observations on the results in table 3 are noteworthy. First, whether or not the data have been collapsed over the covariates, the estimated standard errors of the attributable risk measures under the prospective sampling are always smaller than those under cross-sectional sampling. This is easily explained, noting that under the latter we introduce a second source of variability by estimating the category-specific exposure rates. Second, the estimated PAR (30.1 percent) based on the nominal scale risk model (without a covariate effect) and the same estimated without any model assumption are identical and so are their estimated standard errors. Walter (2) noted that the sum of all estimated category-specific PARs is equal to the estimated marginal PAR, when the marginal PAR is obtained from the 2 × 2 contingency table formed by collapsing over all the exposure categories. The nominal scale logit model without a covariate is in effect a saturated model. Thus, the results in this case are identical to those obtained without any model assumption.

Finally, for these particular data, the covariate-adjusted PAR estimate for BMI (ordinal parametrization) is 36.1 percent, whereas the overall PAR for both BMI and race is only 40.9 percent. As such, these methods provide a promising approach to isolating the relative importance of potentially confounding sources of variation in disease patterns within target populations. Even though the sum of PARs for BMI and race effects does not equal the overall PAR for race and BMI, making it somewhat difficult to quantify the proportion of total variation in DBP explained by BMI, clearly BMI is the dominant source. Benichou (22) suggested that one can obtain an overall PAR estimate from a fully saturated model including the race × BMI interaction terms. In this particular case, the estimated value of this parameter is 36.9 percent, which further shows the influence of BMI over an individual's diastolic blood pressure measures.

DISCUSSION

Model-based estimation leads to increased asymptotic precision when the assumed model adequately describes the data (23). The simpler the model, the better the performance of the model-based estimator. There are several methods of assessing the fit of the model. We have chosen to apply the criterion of p value. If the lack of fit significance level is greater than 5 percent, the model in question is assumed to fit the data. Table 2 shows the significance level for testing the lack of fit of each model using the weighted least squares algorithm as implemented in SAS PROC CATMOD (SAS Institute, Cary, North Carolina).

However, if the model does not fit the data, the model-based asymptotic variances of the estimated parameters are still smaller than their non-model-based counterparts, but the bias of the estimators does not go to 0 with increasing sample size. Misspecification of the model is a part of systematic error in the inference process and requires special attention. In the example considered, we have discussed two models, one model with BMI parametrized as a nominal scale variable and the second model with BMI incorporated as an ordinal variable. There is a gain in efficiency in using Model II, because the asymptotic standard errors for both the prospective and the cross-sectional sampling schemes are smaller than those obtained under Model I. Model II is more parsimonious than Model I, and it still fits the data adequately. However, the gain in efficiency is not overwhelming, and therefore one may reasonably prefer to use Model I. The choice of one model over the other is driven by the relative importance of efficiency versus bias, and, depending on the situation, there may not be a clear winner.

The covariate-adjusted PAR λ_A is equivalent to the formulation in Bruzzi et al. (10). However, Bruzzi et al. considered a case-control study, and under the rare disease assumption they estimated the relative risk by the observed odds ratio. The inference method followed by Greenland and Drescher (18) and the present authors are more general since no such assumption is made.

It has been argued (22) that there is no need to consider more than two levels of exposure to the primary risk factor. Recall that the sum of all estimated category-specific PARs is equal to the estimated marginal PAR when the marginal is obtained by collapsing over the exposed categories (2). This is indeed true in the case of Model I, because it indicates saturation on the primary risk factor. However, under Model II, a trend effect for the primary risk factor is assumed that, when true, may result in substantial savings in the estimated asymptotic variances of the PAR. In this case, the ordinal structure of the risk factor becomes of prime importance and the PAR is not, in general, equivalent to the marginal PAR.

A maximum likelihood estimate of the covariate-adjusted PAR and its asymptotic variance have been derived in this paper utilizing a logit model-based approach. Even though two simplistic models are discussed at length, the methodology is general enough to be adapted to any situation. Currently there is no software available that provides the estimated covariate-adjusted PAR along with its asymptotic variances for cross-sectional study. We used SAS and S-PLUS software (StatSci Division, MathSoft, Inc., Seattle, Washington) to obtain the required estimates.

ACKNOWLEDGMENT

Supported by grant no. PHS 5 R01 HL 33407 from the National Heart, Lung, and Blood Institute, National Institutes of Health.

REFERENCES

1. Levin ML. The occurrence of lung cancer in man. Acta Unio Internationalis Contra Cancrum 1953;9:531–41.
2. Walter SD. The distribution of Levin's measure of attributable risk. Biometrika 1975;62:371–5.
3. Walter SD. Calculation of attributable risks from epidemiological data. Int J Epidemiol 1978;7:175–82.
4. Walter SD. Prevention for multifactorial diseases. Am J Epidemiol 1980;112:409–16.
5. Walter SD. Effects of interaction, confounding and observational error on attributable risk estimation. Am J Epidemiol 1983;117:598–604.
6. Miettinen OS. Proportion of disease caused or prevented by a given exposure, trait or intervention. Am J Epidemiol 1974;99:325–32.
7. Walter SD. The estimation and interpretation of attributable risk in health research. Biometrics 1976;32:829–49.
8. Whittemore AS. Statistical methods for estimating attributable risk from retrospective data. Stat Med 1982;1:229–43.
9. Whittemore AS. Estimating attributable risk from case-control studies. Am J Epidemiol 1983;117:76–85.
10. Bruzzi P, Green SB, Byar DP, et al. Estimating the population attributable risk for multiple risk factors using case-control data. Am J Epidemiol 1985;122:904–14.
11. Benichou J, Gail MH. Variance calculations and confidence intervals for estimates of the attributable risk based on logistic models. Biometrics 1990;46:991–1003.
12. Greenland S. Interpretation and estimation of summary ratios under heterogeneity. Stat Med 1982;1:217–27.
13. Greenland S. Bias in methods for deriving standardized morbidity ratio and attributable fraction estimates. Stat Med 1984;3:131–41.
14. Greenland S. Variance estimators for attributable fraction estimates consistent in both large strata and sparse data. Stat Med 1987;6:701–8.
15. Denman DW, Schlesselman JJ. Interval estimation of the attributable risk for multiple exposure levels in case-control studies. Biometrics 1983;39:185–92.
16. Drescher K, Schill W. Attributable risk estimation from case-control data via logistic regression. Biometrics 1991;47:1247–56.
17. Deubner DC, Wilkinson WE, Helms MJ, et al. Logistic model estimation of death attributable to risk factors for cardiovascular disease in Evans County, Georgia. Am J Epidemiol 1980;112:135–43.
18. Greenland S, Drescher K. Maximum likelihood estimation of the attributable fraction from logistic models. Biometrics 1993;49:865–72.
19. McDowell A, Engel A, Massey JT et al. Plan and operation of the second National Health and Nutrition Examination Survey, 1976–1980. Hyattsville, MD: National Center for Health Statistics, 1981. (Vital and health statistics, Series 1, no. 15) (DHHS publication no. (PHS) 81-1317).
20. Kumanyika SK, Landis JR, Matthews YL, et al. Secular trends in blood pressure among blacks and whites aged 18–34 years in two body mass index strata, United States, 1960–1980. Am J Epidemiol 1994;141:141–54.
21. Benichou J, Gail MH. A delta method for implicitly defined random variables. Am Stat 1989;43:41–4.
22. Benichou J. Methods of adjustment for estimating the attributable risk in case-control studies: a review. Stat Med 1991;10:1753–73.
23. Altham PME. Improving the precision of estimation by fitting a model. J R Stat Soc [B] 1984;46:118–19.