The Distribution of Levin's Measure of Attributable Risk

Author(s): S. D. Walter

Source: *Biometrika*, Vol. 62, No. 2 (Aug., 1975), pp. 371-374
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: https://www.jstor.org/stable/2335374
Accessed: 18-10-2024 20:33 UTC

Biometrika (1975), **62**, 2, *p.* 371
Printed in Great Britain

The distribution of Levin's measure of attributable risk

BY S. D. WALTER
Department of Epidemiology and Community Medicine, University of Ottawa

SUMMARY

Some properties of the sample estimator of attributable risk λ, defined here as the proportion of all cases of disease which may be attributed to a risk factor, are considered for the case-control study situation. It is shown that $\log(1-\hat{\lambda})$ may be expressed in terms of the prevalences of the factor in cases and healthy controls, that the bias of the estimator is minimized when $\frac{1}{2}$ is added to the cell frequencies corresponding to nonexposed persons in the usual 2×2 contingency table, and that the distribution of $1-\hat{\lambda}$ is asymptotically log normal. Examples of the calculations, and a discussion of the results, are given for a number of risk factors for childhood leukemia.

Some key words: Attributable risk; Case-control studies; Childhood leukemia; Contingency table analysis; Multiple risk factors.

1. INTRODUCTION

Suppose that each member of a population may be classified as being either positive or negative for a disease and exposure to an associated risk factor. If the prevalences of disease among individuals exposed or not exposed to the factor are denoted by P_E and P_0 respectively, then the conventional attributable risk is defined as $P_E - P_0$; this quantity represents the excess prevalence among exposed persons that may be attributed to the factor, under the assumption that all other causes generate disease at the same rates in both exposure categories of the particular risk factor under investigation. An alternative measure of attributable risk (Levin, 1953) is defined as the proportion of the totality of disease, i.e. in both exposed and nonexposed persons, which may be attributed to the factor, and this will be denoted by λ. If r is the relative risk and b is the proportion of individuals in the population exposed to the risk factor, then the total disease prevalence is

$$P = bP_E + (1-b)P_0 = brP_0 + (1-b)P_0.$$

The proportion of cases of disease among exposed persons attributable to the factor is $(P_E - P_0)/P_E = (r-1)/r$, so that λ may be derived as

$$\lambda = \frac{brP_0(r-1)/r}{brP_0 + (1-b)P_0} = \frac{b(r-1)}{b(r-1)+1}. \tag{1}$$

Lilienfeld (1973) describes the importance of this statistic, which is also discussed by Cole & MacMahon (1971) in terms of percentages, under the name population attributable risk. This paper considers the estimation and significance of λ when derived from case-control studies.

2. Estimation and significance of λ

Let us suppose that a particular study produces a 2×2 contingency table with cell frequencies as shown in Table 1.

Table 1. *A* 2×2 *sample contingency table*

		Disease		
		+	−	Total
Factor	+	u	v	$u+v$
	−	w	x	$w+x$
Total		$u+w$	$v+x$	$u+v+w+x$

The column totals $n_1 = u+w$ and $n_2 = v+x$ will be presumed constant, appropriately for a case-control study. Then r may be estimated by the usual $\hat{r} = (ux)/(vw)$, and b by $\hat{b} = v/n_2$. Now $\hat{b}$ is a somewhat biased estimator of the proportion of the population with the factor, since b is defined as $Pp_D + (1-P)p_H$, where p_D and p_H are the prevalences of the factor in the diseased and healthy sectors of the population; however it is usually true in practice that $P \ll 1$, in which case $\hat{b}$ is an adequate estimator of b. If $\hat{r}$ and $\hat{b}$ are substituted into (1), we obtain $\hat{\lambda} = (ux - vw)/(xn_1)$. For further development it is more convenient to consider the properties of $\xi = 1-\lambda = 1/\{b(r-1)+1\}$. It may easily be derived that

$$\hat{\xi} = \frac{1}{\hat{b}(\hat{r}-1)+1} = \frac{wn_2}{xn_1}, \tag{2}$$

and hence

$$\log \hat{\xi} = \log(1-\hat{p}_D) - \log(1-\hat{p}_H), \tag{3}$$

where $\hat{p}_D = u/n_1$ and $\hat{p}_H = v/n_2$ are the sample estimates of p_D and p_H; note that $\hat{p}_H = \hat{b}$. Because the case and control groups may usually be regarded as independent samples, we may consider the two logarithms in the right-hand side of (3) separately; in fact this problem is now similar to that for the moments of $\hat{r}$ (Woolf, 1955; Haldane, 1956).

For simplicity we need to consider only $\log(w/n_1)$ as an estimator of $\log q_D = \log(1-p_D)$. The moments of this estimator are infinite because in a finite sample there is in general a nonzero probability of w having the value zero; we must therefore use a modified estimator $\log\{(w+\delta)/(n_1+\delta)\}$, where δ is a positive constant. By considering the binomial moments of w, the bias of the estimator may be derived as

$$E[\log\{(w+\delta)/(n_1+\delta)\} - \log q_D] = \frac{p_D(2\delta-1)}{2n_1 q_D} + \frac{1}{n_1^2}\left\{\frac{p_D}{q_D}(-\delta^2+\delta-\tfrac{1}{3}) + \frac{p_D^2}{q_D^2}(-\tfrac{1}{2}\delta^2+\delta-\tfrac{5}{12})\right\} + O(n_1^{-3}).$$

This provides the standard result that the optimal value of δ is $\frac{1}{2}$, yielding a bias of

$$-p_D(1+q_D)/(24q_D^2 n_1^2) + O(n_1^{-3});$$

in a particular sample this bias would be approximately $-u(n_1+w)/(24w^2n_1^2)$, with a corresponding bias in $\log \hat{\xi}$ approximately equal to

$$\frac{1}{24}\left\{\frac{v(n_2+x)}{x^2n_2^2} - \frac{u(n_1+w)}{w^2n_1^2}\right\}. \tag{4}$$

If $\delta = 0$, that is the possibility of w or x being zero is ignored, the asymptotic bias in $\log \hat{\xi}$ is $\frac{1}{2}\{p_H(n_2 q_H)^{-1} - p_D(n_1 q_D)^{-1}\}$ or approximately $\frac{1}{2}\{v(xn_2)^{-1} - u(wn_1)^{-1}\}$ in a particular sample. Similarly, the second moment of the estimator may be derived from

$$\text{var}\,[\log\{(w+\delta)/(n_1+\delta)\}] = p_D/(n_1 q_D) + O(n_1^{-2}),$$

with a sample equivalent for $\log \hat{\xi}$ as

$$\text{var}\,(\log \hat{\xi}) \doteq \frac{u}{wn_1} + \frac{v}{xn_2}. \tag{5}$$

The higher moments of $\log\{(w+\delta)/(n_1+\delta)\}$ yield a skewness coefficient of

$$(p_D - q_D)^2/(p_D q_D n_1) + O(n_1^{-2})$$

and kurtosis of $3 + O(n_1^{-1})$, and hence the distribution of $\hat{\xi}$ is asymptotically log normal. Approximate confidence limits for ξ may be calculated using the square root of (5) as the standard error of $\log \hat{\xi}$, and applying normal theory to calculate confidence limits for $\log \xi$ which may then be transformed back to give limits for ξ.

3. Numerical example

To illustrate the above calculations, we use some data from the Tri-State Leukemia Study described by Graham *et al.* (1963) and Gibson *et al.* (1968). The project concerns the relationship of the incidence of childhood leukemia with four risk factors, preconception irradiation, reproductive wastage, *in utero* irradiation, and childhood virus disease, and with combinations of these factors. The sample frequencies of cases and controls with and without the risk factors are given in Table 2. The relative risks are calculated after adding $\frac{1}{2}$ to each cell frequency as suggested by Haldane (1956); for example, the relative risk for preconception irradiation is estimated as $(101{\cdot}5 \times 114{\cdot}5)/(68{\cdot}5 \times 122{\cdot}5) = 1{\cdot}38$. The parameter b is estimated from the control sample as $122/236 = 0{\cdot}52$ for this first factor. After addition of $\frac{1}{2}$ to w, x, n_1 and n_2, we derive $\hat{\xi}$ from (2) as $(68{\cdot}5 \times 236{\cdot}5)/(114{\cdot}5 \times 169{\cdot}5) = 0{\cdot}835$; hence the estimate of λ is $1 - 0{\cdot}835 = 0{\cdot}165$. The bias in $\log \hat{\xi}$, given by (4), may be evaluated as $-5{\cdot}09 \times 10^{-6}$, effectively negligible. To evaluate the significance of the departure of $\log \hat{\xi}$ from zero, an exact test of association, or any of the usual approximations to it, may be performed on the original frequencies in the 2×2 table. For a large-sample test on $\log \hat{\xi}$, we have $\log \hat{\xi} = -0{\cdot}1806$ and from (5) that $\text{var}\,(\log \hat{\xi}) \doteq 0{\cdot}0133$. If we suppose that the distribution of $\log \hat{\xi}$ is approximately normal, the test statistic may be calculated as $-0{\cdot}1806/\sqrt{0{\cdot}0133} = -1{\cdot}57$, which is not significant at the 5 % level. The figures for the other factors are calculated similarly. Slight differences may be noted in the relative and attributable risks between Table 2 and the corresponding figures in Table 3 of Lilienfeld (1973); these are due to the use of an adjustment δ in the present analysis, with the exception of the attributable risk in the cumulative factor analysis, where a numerical error is present in the original calculations.

Some simulation experiments were carried out to generate the null distributions of $\hat{r}$ and $\log \hat{\xi}$ for various sets of values for n_1, n_2 and b, including those corresponding to the data for each risk factor given in Table 2. It was found that both distributions had very small skewness, but were somewhat platykurtic. The significance levels calculated using normal theory are therefore somewhat conservative. The differences between the significance levels calculated from normal theory and empirically from the simulation results were somewhat greater for the attributable risk, and when the proportion b was close to

either zero or one. However, in practice, the normal theory results seem adequate for values of b between 0·1 and 0·9 for the sample sizes considered here.

It may be noted that although reproductive wastage and *in utero* irradiation appear to be significant risk factors, the largest attributable risk is only 25 %, for the cumulative factor category. This may imply that there is at least one other important causative factor involved which was not included in the study, but which may be partially confounded with the four factors discussed here. The nonsignificance of the attributable risk for the cumulative category may be explained by the rather small number of cases and controls not exposed to any factor, yielding a fairly high variance for the risk. The apparent effects of all the factors studied may also have been reduced somewhat by imperfect recording of the relevant events.

Table 2. *Relative and attributable risks for leukemia in children (1–4 years) for four risk factors*

Risk factor	No. of cases		No. of controls		$\hat{r}$	$\hat{b}$	$\hat{\lambda}$	
	+	−	+	−				
Preconception irradiation	101	68	122	114	1·38	0·52	0·165	N.S.
Reproductive wastage	48	122	42	196	1·83	0·18	0·128	**
In utero irradiation	64	105	65	171	1·60	0·28	0·142	*
Childhood virus diseases	57	96	70	156	1·32	0·31	0·090	N.S.
Cumulative (1 or more factors)	130	21	181	42	1·42	0·81	0·254	N.S.

N.S. $P \geqslant 0{\cdot}05$; *$P < 0{\cdot}05$; **$P < 0{\cdot}01$ (one-sided tests).
+, with factor; −, without factor.

Although for a given set of data, significance tests on the relative and attributable risks are in principle identical, Levin's measure of attributable risk does have the important practical advantage in that it will provide an estimate of the expected reduction in the number of cases which would occur if the risk factor under study were removed. Thus it might be used, albeit rather crudely, to establish policy priorities in the strategy of disease eradication and preventive medicine.

I am grateful to Professor Lilienfeld of the Johns Hopkins University and to a referee for some helpful suggestions in the development of this paper. The author is supported by the Ontario Ministry of Health.

REFERENCES

COLE, P. & MACMAHON, B. (1971). Attributable risk percent in case-control studies. *Brit. J. Prev. Soc. Med.* **25**, 242–4.

GIBSON, R. W., BROSS, I. D. J., GRAHAM, S., LILIENFELD, A. M., SCHUMAN, L. M., LEVIN, M. L. & DOWD, J. E. (1968). Leukemia in children exposed to multiple risk factors. *N. Eng. J. Med.* **279**, 906–9.

GRAHAM, S., LEVIN, M. L., LILIENFELD, A. M., DOWD, J. E., SCHUMAN, L. M., GIBSON, R., HEMPELMANN, L. H. & GERHARDT, P. (1963). Methodological problems and design of the Tri-State Leukemia Study. *Ann. N.Y. Acad. Sci.* **107**, 557–69.

HALDANE, J. B. S. (1956). Estimation and significance of the logarithm of a ratio of frequencies. *Ann. Hum. Gen.* **20**, 309–11.

LEVIN, M. L. (1953). The occurrence of lung cancer in man. *Acta Un. Intern. Cancer* **9**, 531–41.

LILIENFELD, A. M. (1973). Epidemiology of infectious and non-infectious disease: some comparisons. *Am. J. Epid.* **97**, 135–47.

WOOLF, B. (1955). On estimating the relation between blood group and disease. *Ann. Hum. Gen.* **19**, 251–3.

[*Received May* 1974. *Revised November* 1974]