Attributable Risk for a Multilevel Risk Factor at Intermediate Base Levels Under a Cross-sectional Study

Khairul Islam[1] and Tanweer Shapla[2]
[1,2]Department of Mathematics, Eastern Michigan University, Ypsilanti, MI 48197 USA

Abstract

Attributable risk is one of the most popular indices for measuring the risk of a factor in the development of disease in epidemiology and biomedical science. The usual practice for a multilevel risk factor is to determine the category specific attributable risk at different levels reference to a base level. When the risk factor is significant, it is also important to evaluate the risk of the factor at different levels reference to intermediate base levels. This information would contribute policy makers and health practitioners in detecting important categorization of the risk factor at a multilevel setting. This paper investigates confidence interval estimate of attributable risk for a multilevel exposure factor using intermediate base levels under a cross-sectional study and provides an application using a real-life example.

Key Words: Attributable Risk, Cross-sectional Study, Confidence Intervals, Delta Method.

1. Introduction

The attributable risk (AR) is one of the most important epidemiological indices for assessing the potential impact of a risk factor and comparing various prevention strategies. It has been used by the epidemiologists and public health administrators to locate the factors that may increase the chance of developing a particular disease and take initiatives to prevent those factors. Introduced by Levin (1953), it is defined as the proportion of the disease that could be avoided if the risk factor were totally eliminated from the population of interest. It has also been termed as etiologic fraction and fraction of etiology (Miettinen, 1974), attributable fraction (Ouellet et al., 1979; Greenland and Robins, 1988; Last, 1983), and population attributable risk per cent (Cole and MacMahon, 1971). A number of articles have dealt with methodologies for the estimation of AR in cross-sectional studies for a dichotomous risk factor (Walter, 1976; Fleiss, 1979; Lui, 2001). In real life, there are situations when concentration on an exposure variable with multiple levels is more appealing. For example, because smoking has a relatively high exposed rate in the population, one might be interested in the estimation of level or category specific risk of smoking on the development of a certain disease. The level-specific AR corresponds to the proportion of disease cases that could be attributed to a specified level of exposure, and may play an important role in disease prevention strategies (Denman and Schlesselman, 1983; Miettinen, 1974; Walter, 1976). The estimation of level-specific AR has been studied by Lui (2003) for case-control studies considering the baseline or reference level with that of the lowest disease risk.

Correspondence: mislam4@emich.edu

This paper concentrates on the estimation of AR for a risk factor having multiple levels for a cross-sectional study design in two different approaches: AR with respect to the baseline or reference level and AR with respect to intermediate level. The idea of AR with respect to intermediate level may help reduce the number of levels of the risk factor if this particular level turns out to be insignificant. This way one can avoid unnecessary categorization of a risk factor. In addition to this, we can get the proportion of disease reduction between the intermediate levels, rather than only considering the baseline or reference level. This information might be useful in epidemiologic study in targeting level specific risk reduction and hence in planning appropriate disease prevention programs.

Section 2 of this paper reviews the statistical model to estimate AR in a cross-sectional study design. In Section 2.1, we use delta method (Agresti, 2002) to derive the asymptotic variance of the estimate of AR. In Section 2.2, we develop the expression for AR_{ij}, the attributable risk for reducing level j to level i. As an application, a real life example has been provided in Section 3 to illustrate the method where the disease of interest is hypertension and BMI has been considered as a risk factor.

2. Statistical Development

Given a risk factor with $J+1$ exposure levels designated by j, $j = 0, 1, ..., J$, we wish to estimate AR_{ij}, the attributable risk for reducing the exposure from level j to i under a cross-sectional study design. Given a random sample of size n, cross-classify each individual according to the status of the disease outcome variable D designated by *k*, *k*=0, 1where 0(1) means the absence (presence) of the disease. Let n_{jk} be the random frequency of n individuals falling into the cell with exposure level j and disease status k and let π_{jk} be the corresponding probability. The Table 1 below summarizes the data structure for a cross-sectional design with a multilevel risk factor and a dichotomous disease outcome variable.

Table 1: Cross-classifying n subjects according to their exposure levels and the status of the disease

Exposure Levels	Disease Status, D		
	Absent (0)	Present (1)	Total
0	n_{00} (π_{00})	n_{01} (π_{01})	$n_{0.}(\pi_{0.})$
1	n_{10} (π_{10})	n_{11} (π_{11})	$n_{1.}(\pi_{1.})$
.	.	.	.
.	.	.	.
J	n_{J0} (π_{J0})	n_{J1} (π_{J1})	$n_{J.}$ $(\pi_{J.})$
Total	$n_{.0}$ $(\pi_{.0})$	$n_{.1}(\pi_{.1})$	$n(1)$

Then the random vector $\mathbf{N}$ given by $\mathbf{N}' = (n_{00}, n_{01}, n_{10}, n_{11}, ..., n_{J0}, n_{J1})$ follows the multinomial distribution with parameters n and $\boldsymbol{\pi}$ given by $\boldsymbol{\pi}' = (\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11}, ..., \pi_{J0}, \pi_{J1})$. The maximum likelihood estimator (MLE) of $\boldsymbol{\pi}$, $\mathbf{p}$ is given by $\mathbf{p}' = (p_{00}, p_{01}, p_{10}, p_{11}, ..., p_{J0}, p_{J1})$ where $p_{jk} = \frac{n_{jk}}{n}$ is the MLE of π_{jk}. When the number of subjects, n, is large, by the Multivariate Central Limit Theorem (Rao, 1973), the random vector $(\mathbf{p} - \boldsymbol{\pi})$ is asymptotically distributed as normal $N(\mathbf{0}, \boldsymbol{\Sigma})$, where $\mathbf{0}' = (0, 0, ..., 0)$ and $\boldsymbol{\Sigma}$ is $2(J+1) \times 2(J+1)$ covariance matrix of the estimate $\mathbf{p}$ of $\boldsymbol{\pi}$ given by

$$\Sigma = \frac{1}{n}\begin{bmatrix} \pi_{00}(1-\pi_{00}) & -\pi_{00}\pi_{01} & -\pi_{00}\pi_{10} & -\pi_{00}\pi_{11} \ldots -\pi_{00}\pi_{J0} & -\pi_{00}\pi_{J1} \\ -\pi_{01}\pi_{00} & \pi_{01}(1-\pi_{01}) & -\pi_{01}\pi_{10} & -\pi_{01}\pi_{11} \ldots -\pi_{01}\pi_{J0} & -\pi_{01}\pi_{J1} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ -\pi_{J1}\pi_{00} & -\pi_{J1}\pi_{01} & -\pi_{J1}\pi_{10} & -\pi_{J1}\pi_{11} \ldots -\pi_{J1}p_{J1} & \pi_{J1}(1-\pi_{J1}) \end{bmatrix}$$

The AR of a disease for reducing the exposure from level j to 0 is defined by

$$\mathrm{AR}_j = \frac{[P(D=1 \mid E=j) - P(D=1 \mid E=0)]P(E=j)}{P(D)} \tag{2.1}$$

Equation (2.1) can be written as

$$\mathrm{AR}_j = P(E=j \mid D=1) - \frac{P(E=j \mid D=1)}{\mathrm{RR}_j} \tag{2.2}$$

where $\mathrm{RR}_j = P(D=1 \mid E=j) / P(D=1 \mid E=0)$ is the relative risk between exposure levels j and 0.

By the definition of conditional probability, it is easy to see that $P(E=j \mid D=1) = \frac{\pi_{j1}}{\pi_{.1}}$.

Also by definition, $\mathrm{RR}_j = \frac{P(D=1 \mid E=j)}{P(D=1 \mid E=0)} = \frac{\pi_{j1}/\pi_{j.}}{\pi_{01}/\pi_{0.}} = \frac{\pi_{j1}\,\pi_{0.}}{\pi_{01}\,\pi_{j.}}$

Then it follows from equation (2.2)

$$\mathrm{AR}_j = \frac{\pi_{j1}}{\pi_{.1}} - \frac{\pi_{j1}}{\pi_{.1}} \frac{\pi_{01}\,\pi_{j.}}{\pi_{j1}\,\pi_{0.}} = \frac{1}{\pi_{.1}}\left\{\pi_{j1} - \frac{\pi_{01}\,\pi_{j.}}{\pi_{0.}}\right\} \tag{2.3}$$

By the invariance property of the MLE, the MLE of AR_j, $\hat{\mathrm{AR}}_j$, is given by $\hat{\mathrm{AR}}_j = \frac{1}{p_{.1}}\left\{p_{j1} - \frac{p_{01}\,p_{j.}}{p_{0.}}\right\}$.

2.1 Derivation of the Asymptotic Variance of $\hat{\mathrm{AR}}_j$

Let $\boldsymbol{\varphi}$ be the vector of partial derivatives of $\hat{\mathrm{AR}}_j$ with respect to the components of the vector $\mathbf{p}$ evaluated at $\mathbf{p} = \boldsymbol{\pi}$. Then we have,

$$\boldsymbol{\varphi}' = \left(\frac{\partial \mathrm{AR}_j}{\partial \pi_{00}}, \frac{\partial \mathrm{AR}_j}{\partial \pi_{01}}, \frac{\partial \mathrm{AR}_j}{\partial \pi_{10}}, \frac{\partial \mathrm{AR}_j}{\partial \pi_{11}}, \ldots, \frac{\partial \mathrm{AR}_j}{\partial \pi_{J0}}, \frac{\partial \mathrm{AR}_j}{\partial \pi_{J1}}\right)$$

where,

$$\frac{\partial \mathrm{AR}_j}{\partial \pi_{00}} = -\frac{1}{{\pi_{.0}}^2}\left\{\pi_{j0} - \frac{\pi_{00}\,\pi_{j.}}{\pi_{0.}}\right\} + \frac{1}{\pi_{.0}}\left\{0 - \left(\frac{\pi_{0.}\,\pi_{j.} - \pi_{00}\,\pi_{j.}}{{\pi_{0.}}^2}\right)\right\}$$

$$= -\frac{1}{\pi_{.0}}\left(\mathrm{AR}_j + \frac{\pi_{01}\,\pi_{j.}}{{\pi_{0.}}^2}\right)$$

$$\frac{\partial \mathrm{AR}_j}{\partial \pi_{m0}} = -\frac{1}{{\pi_{.0}}^2}\left\{\pi_{j0} - \frac{\pi_{00}\,\pi_{j.}}{\pi_{0.}}\right\} = -\frac{1}{\pi_{.0}}\mathrm{AR}_j, \text{ for } m \neq j, m \geq 1$$

$$\frac{\partial \mathrm{AR}_j}{\partial \pi_{j0}} = -\frac{1}{{\pi_{.0}}^2}\left\{\pi_{j0} - \frac{\pi_{00}\,\pi_{j.}}{\pi_{0.}}\right\} + \frac{1}{\pi_{.0}}\left\{1 - \frac{\pi_{00}}{\pi_{0.}}\right\}$$

$$= -\frac{1}{\pi_{.0}}(\mathrm{AR}_j - \frac{\pi_{01}}{\pi_{0.}}), \text{since } \pi_{0.} = \pi_{00} + \pi_{01}$$

$$\frac{\partial \mathrm{AR}_j}{\partial \pi_{01}} = \frac{1}{\pi_{.0}}\left\{\frac{\pi_{00}\,\pi_{j.}}{{\pi_{0.}}^2}\right\} = \frac{\pi_{00}\,\pi_{j.}}{\pi_{.0}\,{\pi_{0.}}^2}$$

$$\frac{\partial \mathrm{AR}_j}{\partial \pi_{m1}} = 0, \text{for } m \neq j, m \geq 1$$

$$\frac{\partial \mathrm{AR}_j}{\partial \pi_{j1}} = \frac{1}{\pi_{.1}}\left\{0 - \frac{\pi_{00}}{\pi_{0.}}\right\} = -\frac{\pi_{00}}{\pi_{.0}\,\pi_{0.}}$$

By using the delta method, the asymptotic variance of $\hat{\mathrm{AR}}_j, V(\hat{\mathrm{AR}}_j)$ is $\boldsymbol{\varphi}'\boldsymbol{\Sigma}\boldsymbol{\varphi}$.

Lemma 2.1 *Under above notations*

$$\boldsymbol{\varphi}'\boldsymbol{\Sigma}\boldsymbol{\varphi} = \frac{1}{n}\left\{\sum_i\sum_k \pi_{ik}\left(\frac{\partial \mathrm{AR}_j}{\partial \pi_{ik}}\right)^2 - {A_{ik}}^2\right\} \text{ where } A_{ik} = \sum_i\sum_k \pi_{ik}\left(\frac{\partial \mathrm{AR}_j}{\partial \pi_{ik}}\right).$$

Proof: The variance covariance matrix $\boldsymbol{\Sigma}$ of the vector $\mathbf{p}$ can be expressed as

$$\boldsymbol{\Sigma} = \frac{1}{n}[diag(\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11}, ..., \pi_{J0}, \pi_{J1}) - \boldsymbol{\pi}\boldsymbol{\pi}']$$

where,

$$diag(\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11}, ..., \pi_{J0}, \pi_{J1}) = \begin{bmatrix} \pi_{00} & & & & & & \\ & \pi_{01} & & & & & \\ & & \pi_{10} & & & & \\ & \mathbf{0} & & \pi_{11} & & \mathbf{0} & \\ & & & & \ddots & & \\ & & & & & \pi_{J0} & \\ & & & & & & \pi_{J1} \end{bmatrix}$$

and $\boldsymbol{\pi}' = (\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11}, ..., \pi_{J0}, \pi_{J1})$.

Then, $n\boldsymbol{\varphi}'\boldsymbol{\Sigma}\boldsymbol{\varphi} = \boldsymbol{\varphi}' diag(\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11}, ..., \pi_{J0}, \pi_{J1})\boldsymbol{\varphi} - \boldsymbol{\varphi}'\boldsymbol{\pi}\boldsymbol{\pi}'\boldsymbol{\varphi}$. After simplification, it follows that,

$$\boldsymbol{\varphi}' diag(\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11}, ..., \pi_{J0}, \pi_{J1})\boldsymbol{\varphi} = \sum_i \sum_k \pi_{ik} \left(\frac{\partial \mathrm{AR}_j}{\partial \pi_{ik}} \right)^2$$

and $\boldsymbol{\varphi}'\boldsymbol{\pi}\boldsymbol{\pi}'\boldsymbol{\varphi} = \left[\sum_i \sum_k \pi_{ik} \left(\frac{\partial \mathrm{AR}_j}{\partial \pi_{ik}} \right) \right]^2 = {A_{ik}}^2$.

Hence the lemma follows.

Theorem 2.2 *The asymptotic variance of* $\hat{\mathrm{AR}}_j, V(\hat{\mathrm{AR}}_j)$ *using the delta method is given by*

$$V(\hat{\mathrm{AR}}_j) = \frac{1}{n\,{\pi_{.0}}^2} \left\{ \begin{array}{l} {\mathrm{AR}_j}^2 \pi_{.0} + 2\,\mathrm{AR}_j \dfrac{\pi_{01}(\pi_{00}\,\pi_{j.} - \pi_{j0}\,\pi_{0.})}{{\pi_{0.}}^2} + \dfrac{\pi_{j0}\,{\pi_{01}}^2 + \pi_{j1}\,{\pi_{00}}^2}{{\pi_{0.}}^2} \\ + \dfrac{\pi_{00}\,{\pi_{j.}}^2\,\pi_{01}}{{\pi_{0.}}^3} \end{array} \right\}$$

Proof: Note that,

$$\sum_i \sum_k \pi_{ik} \left(\frac{\partial \mathrm{AR}_j}{\partial \pi_{ik}} \right)^2$$

$$= \pi_{00} \left(\frac{\partial \mathrm{AR}_j}{\partial \pi_{00}} \right)^2 + \pi_{j0} \left(\frac{\partial \mathrm{AR}_j}{\partial \pi_{j0}} \right)^2 + \sum_{i=1, i\neq j}^{J} \pi_{i0} \left(\frac{\partial \mathrm{AR}_j}{\partial \pi_{i0}} \right)^2 + \pi_{01} \left(\frac{\partial \mathrm{AR}_j}{\partial \pi_{01}} \right)^2 + \pi_{j1} \left(\frac{\partial \mathrm{AR}_j}{\partial \pi_{j1}} \right)^2$$

$$+ \sum_{i=1, i\neq j}^{J} \pi_{i1} \left(\frac{\partial \mathrm{AR}_j}{\partial \pi_{i1}} \right)^2$$

$$= \pi_{00} \left\{ \frac{1}{{\pi_{.0}}^2} \left(\mathrm{AR}_j + \frac{\pi_{01}\,\pi_{j.}}{{\pi_{0.}}^2} \right)^2 \right\} + \pi_{j0} \left\{ \frac{1}{{\pi_{.0}}^2} \left(\mathrm{AR}_j - \frac{\pi_{01}}{\pi_{0.}} \right)^2 \right\} + \sum_{i=1, i\neq j}^{J} \pi_{i0} \left\{ \frac{1}{{\pi_{.0}}^2} {\mathrm{AR}_j}^2 \right\}$$

$$+ \pi_{01} \left(\frac{\pi_{00}\,\pi_{j.}}{\pi_{.0}\,{\pi_{0.}}^2} \right)^2 + \pi_{j1} \left(-\frac{\pi_{00}}{\pi_{.0}\,\pi_{0.}} \right)^2$$

$$= \frac{{\mathrm{AR}_j}^2 (\pi_{00} + \pi_{j0} + \sum_{i=1, i\neq j}^{J} \pi_{i0})}{{\pi_{.0}}^2} + \frac{2\,\mathrm{AR}_j}{{\pi_{.0}}^2} \frac{\pi_{00}\,\pi_{j.}\,\pi_{01}}{{\pi_{0.}}^2} + \frac{1}{{\pi_{.0}}^2} \frac{\pi_{00}\,{\pi_{j.}}^2\,{\pi_{01}}^2}{{\pi_{0.}}^4} - \frac{2\,\mathrm{AR}_j}{{\pi_{.0}}^2} \frac{\pi_{j0}\,\pi_{01}}{\pi_{0.}}$$

$$+ \frac{\pi_{j0}\,{\pi_{01}}^2}{{\pi_{.0}}^2 {\pi_{0.}}^2} + \frac{{\pi_{00}}^2\,{\pi_{j.}}^2\,\pi_{01}}{{\pi_{.0}}^2\,{\pi_{0.}}^4} + \frac{1}{{\pi_{.0}}^2} \frac{\pi_{j1}\,{\pi_{00}}^2}{{\pi_{0.}}^2}$$

$$= \frac{1}{{\pi_{.0}}^2} \left\{ {\mathrm{AR}_j}^2 \pi_{.0} + 2\,\mathrm{AR}_j \frac{\pi_{01}(\pi_{00}\,\pi_{j.} - \pi_{j0}\,\pi_{0.})}{{\pi_{0.}}^2} + \frac{\pi_{j0}\,{\pi_{01}}^2 + \pi_{j1}\,{\pi_{00}}^2}{{\pi_{0.}}^2} + \frac{\pi_{00}\,{\pi_{j.}}^2\,\pi_{01}}{{\pi_{0.}}^3} \right\}$$

Also,

$$A_{ik} = \sum_i \sum_k \pi_{ik} \left(\frac{\partial \mathrm{AR}_j}{\partial \pi_{ik}} \right)$$

$$= \pi_{00}\left(\frac{\partial \mathrm{AR}_j}{\partial \pi_{00}}\right) + \pi_{j0}\left(\frac{\partial \mathrm{AR}_j}{\partial \pi_{j0}}\right) + \sum_{i=1, i\neq j}^{J} \pi_{i0}\left(\frac{\partial \mathrm{AR}_j}{\partial \pi_{i0}}\right) + \pi_{01}\left(\frac{\partial \mathrm{AR}_j}{\partial \pi_{01}}\right) + \pi_{j1}\left(\frac{\partial \mathrm{AR}_j}{\partial \pi_{j1}}\right)$$

$$+ \sum_{i=1, i\neq j}^{J} \pi_{i1}\left(\frac{\partial \mathrm{AR}_j}{\partial \pi_{i1}}\right)$$

$$= \pi_{00}\left\{-\frac{1}{\pi_{.0}}\left(\mathrm{AR}_j + \frac{\pi_{01}\pi_{j.}}{\pi_{0.}^{2}}\right)\right\} + \pi_{j0}\left\{-\frac{1}{\pi_{.0}}\left(\mathrm{AR}_j - \frac{\pi_{01}}{\pi_{0.}}\right)\right\} + \sum_{i=1, i\neq j}^{J} \pi_{i0}\left\{-\frac{1}{\pi_{.0}}\mathrm{AR}_j\right\}$$

$$+ \pi_{01}\left(\frac{\pi_{00}\pi_{j.}}{\pi_{.0}\pi_{0.}^{2}}\right) + \pi_{j1}\left(-\frac{\pi_{00}}{\pi_{.0}\pi_{0.}}\right)$$

$$= -\frac{\mathrm{AR}_j}{\pi_{.0}}(\pi_{00} + \pi_{j0} + \sum_{i=1, i\neq j}^{J}\pi_{i0}) - \frac{\pi_{00}(\pi_{0.} - \pi_{00})\pi_{j.}}{\pi_{.0}\pi_{0.}^{2}} + \frac{\pi_{j0}\pi_{01}}{\pi_{.0}\pi_{0.}} + \frac{\pi_{00}\pi_{01}\pi_{j.}}{\pi_{.0}\pi_{0.}^{2}} - \frac{\pi_{j1}\pi_{00}}{\pi_{.0}\pi_{0.}}$$

$$= -\frac{1}{\pi_{.0}}\left(\mathrm{AR}_j\,\pi_{.0} + \frac{\pi_{00}\pi_{01}\pi_{j.}}{\pi_{0.}^{2}} - \frac{\pi_{j0}\pi_{01}}{\pi_{0.}} - \frac{\pi_{00}\pi_{01}\pi_{j.}}{\pi_{0.}^{2}} + \frac{\pi_{j1}\pi_{00}}{\pi_{0.}}\right)$$

$$= -\frac{1}{\pi_{.0}}\left(\mathrm{AR}_j\,\pi_{.0} + \frac{\pi_{j1}\pi_{00} - \pi_{j0}\pi_{01}}{\pi_{0.}}\right)$$

$$= -\frac{1}{\pi_{.0}}\left(\mathrm{AR}_j\,\pi_{.0} - \mathrm{AR}_j\,\pi_{.0}\right)$$

$$= 0$$

Therefore, the asymptotic variance of $\hat{\mathrm{AR}}_j, V(\hat{\mathrm{AR}}_j)$, follows immediately.

An estimate of $V(\hat{\mathrm{AR}}_j)$, $\hat{V}(\hat{\mathrm{AR}}_j)$, can be obtained by substituting the MLEs p_{jk} for π_{jk} and $\hat{\mathrm{AR}}_j$ for AR_j. Then an asymptotic $100\ (1-\alpha)$ percent confidence interval for AR_j using Wald's statistic is given by

$$\left[\hat{\mathrm{AR}}_j - z_{\alpha/2}\sqrt{\hat{V}(\hat{\mathrm{AR}}_j)}, \min\left(\hat{\mathrm{AR}}_j + z_{\alpha/2}\sqrt{\hat{V}(\hat{\mathrm{AR}}_j)}, 1\right)\right],$$

where z_α is the upper $100\ (\alpha)$ th percentile of the standard normal distribution.

2.2 Attributable Risk at Intermediate Base-level

In this section, we define intermediate base-level attributable risk at exposure level j with intermediate base level i, denoted by AR_{ij}. The intermediate base-level attributable risk, AR_{ij}, can be expressed in terms of AR_i and AR_j, $0 < i < j$.

The intermediate base-level AR of a disease for reducing the exposure from level j to i, is defined by

$$\mathrm{AR}_{ij} = \frac{[P(D=1|E=j) - P(D=1|E=i)]P(E=j)}{P(D)}$$

Using the cell probabilities, it is easy to see that

$$\mathrm{AR}_{ij} = \frac{1}{\pi_{.0}}\left\{\pi_{j0} - \frac{\pi_{i0}\,\pi_{j.}}{\pi_{i.}}\right\} \tag{2.4}$$

Theorem 2.3: *The intermediate base-level attributable risk* AR_{ij} *for reducing exposure from level* j *to* i *can be expressed as the base-level attributable risks* AR_j *and* AR_i *as follows*

$$\mathrm{AR}_{ij} = \mathrm{AR}_j - \mathrm{AR}_i \frac{\pi_{j.}}{\pi_{i.}}$$

Proof: Following equation (2.3), the AR of a disease for reducing the exposure from level i to 0 we have,

$$\mathrm{AR}_i = \frac{1}{\pi_{.0}}\left\{\pi_{i0} - \frac{\pi_{00}\,\pi_{i.}}{\pi_{0.}}\right\} \tag{2.5}$$

In order to express AR_{ij} in terms of AR_j and AR_i, let us write

$$\mathrm{AR}_{ij} = \mathrm{AR}_j - x\mathrm{AR}_i .$$

This implies that $x = \dfrac{\mathrm{AR}_j - \mathrm{AR}_{ij}}{\mathrm{AR}_i}$.

Then, by equations (2.3), (2.4) and (2.5) we have,

$$x = \frac{\frac{1}{\pi_{.0}}\left\{\pi_{j0} - \frac{\pi_{00}\,\pi_{j.}}{\pi_{0.}}\right\} - \frac{1}{\pi_{.0}}\left\{\pi_{j0} - \frac{\pi_{i0}\,\pi_{j.}}{\pi_{i.}}\right\}}{\frac{1}{\pi_{.0}}\left\{\pi_{i0} - \frac{\pi_{00}\,\pi_{i.}}{\pi_{0.}}\right\}}$$

$$= \frac{\frac{1}{\pi_{.0}}\left\{\pi_{j0} - \pi_{j0} + \pi_{j.}\left(\frac{\pi_{i0}}{\pi_{i.}} - \frac{\pi_{00}}{\pi_{0.}}\right)\right\}}{\frac{1}{\pi_{.0}}\left\{\pi_{i0} - \frac{\pi_{00}\,\pi_{i.}}{\pi_{0.}}\right\}}$$

$$= \frac{\pi_{j.}\left(\pi_{i0}\,\pi_{0.} - \pi_{00}\pi_{i.}\right)/\,\pi_{0.}\pi_{i.}}{\left(\pi_{i0}\pi_{0.} - \pi_{00}\,\pi_{i.}\right)/\,\pi_{0.}}$$

$$= \frac{\pi_{j.}}{\pi_{i.}}$$

Hence the theorem follows.

By the invariance property of the MLE, the MLE of AR_{ij}, $\hat{\mathrm{AR}}_{ij}$, is given by

$$\hat{\mathrm{AR}}_{ij} = \hat{\mathrm{AR}}_j - \hat{\mathrm{AR}}_i \frac{p_{j.}}{p_{i.}} .$$

Theorem 2.4: *An asymptotic variance of* $\hat{\mathrm{AR}}_{ij}$ *can be expressed as the base-level attributable risks* AR_j *and* AR_i *as follows*

$$V(\hat{\mathrm{AR}}_{ij})) = \frac{1}{n\,\pi_{.0}^{\,2}} \left\{ \begin{array}{l} \dfrac{\pi_{.0}(\pi_{i0} - \pi_{i1})}{\pi_{i.}} \mathrm{AR}_j^{\,2} + 2\,\mathrm{AR}_i\,\mathrm{AR}_j \dfrac{\pi_{i1}\pi_{j.}\pi_{.0}}{\pi_{i.}^{\,2}} + \dfrac{\pi_{j0}\,\pi_{i1}^{\,2} + \pi_{j1}\,\pi_{i0}^{\,2}}{\pi_{i.}^{\,2}} \\ + \dfrac{\pi_{i0}\,\pi_{j.}^{\,2}\,\pi_{i1}}{\pi_{i.}^{\,3}} \end{array} \right\}$$

Proof: The variance of the estimator $\hat{\mathrm{AR}}_{ij}$ can be obtained by replacing exposure level 0 by i in the expression of $V(\hat{\mathrm{AR}}_j)$ and is given by

$$V(\hat{\mathrm{AR}}_{ij})) = \frac{1}{n\,\pi_{.0}^{\,2}} \left\{ \begin{array}{l} \mathrm{AR}_j^{\,2}\pi_{.0} + 2\,\mathrm{AR}_j \dfrac{\pi_{i1}(\pi_{i0}\,\pi_{j.} - \pi_{j0}\,\pi_{i.})}{\pi_{i.}^{\,2}} + \dfrac{\pi_{j0}\,\pi_{i1}^{\,2} + \pi_{j1}\,\pi_{i0}^{\,2}}{\pi_{i.}^{\,2}} \\ + \dfrac{\pi_{i0}\,\pi_{j.}^{\,2}\,\pi_{i1}}{\pi_{i.}^{\,3}} \end{array} \right\} \quad (2.6)$$

From (2.4) we have,

$$\pi_{j0}\,\pi_{i.} - \pi_{i0}\,\pi_{j.} = \pi_{.0}\,\pi_{i.}\mathrm{AR}_{ij}$$
$$\Rightarrow \pi_{i0}\,\pi_{j.} - \pi_{j0}\,\pi_{i.} = -\pi_{.0}\,\pi_{i.}\mathrm{AR}_{ij}$$

Because $\mathrm{AR}_{ij} = \mathrm{AR}_j - \mathrm{AR}_i \dfrac{\pi_{j.}}{\pi_{i.}}$, we have

$$\pi_{i0}\,\pi_{j.} - \pi_{j0}\,\pi_{i.} = -\pi_{.0}\,\pi_{i.}\left(\mathrm{AR}_j - \mathrm{AR}_i \frac{\pi_{j.}}{\pi_{i.}}\right)$$
$$= -\pi_{.0}\,\pi_{i.}\mathrm{AR}_j + \mathrm{AR}_i\pi_{.0}\pi_{j.}$$

Then from the first two terms on the right hand side of the expression of $V(\hat{\mathrm{AR}}_{ij})$ in (2.6) within the braces

$$\mathrm{AR}_j^{\,2}\pi_{.0} + 2\mathrm{AR}_j \frac{\pi_{i1}(-\pi_{.0}\,\pi_{i.}\mathrm{AR}_j + \mathrm{AR}_i\pi_{.0}\pi_{j.})}{\pi_{i.}^{\,2}}$$

$$= \mathrm{AR}_j^{\,2}\pi_{.0} - 2\mathrm{AR}_j^{\,2}\frac{\pi_{i1}\pi_{.0}}{\pi_{i.}} + 2\mathrm{AR}_i\,\mathrm{AR}_j\frac{\pi_{i1}\pi_{.0}\pi_{j.}}{\pi_{i.}^{\,2}}$$

$$= \frac{(\pi_{i.} - 2\pi_{i1})\pi_{.0}}{\pi_{i.}}\mathrm{AR}_j^{\,2} + 2\mathrm{AR}_i\,\mathrm{AR}_j\frac{\pi_{i1}\pi_{.0}\pi_{j.}}{\pi_{i.}^{\,2}}$$

$$= \frac{(\pi_{i0} - \pi_{i1})\pi_{.0}}{\pi_{i.}}\mathrm{AR}_j^{\,2} + 2\mathrm{AR}_i\,\mathrm{AR}_j\frac{\pi_{i1}\pi_{.0}\pi_{j.}}{\pi_{i.}^{\,2}}, \text{ because } \pi_{i.} = \pi_{i0} + \pi_{i1}$$

Hence the result follows immediately.

3. Application

As an application, here we provide an example using data from a sample of 966 subjects obtained from the Second National Health and Nutrition Examination Survey (NHANES

II) conducted from 1976 to 1980 (McDowell et al., 1981). The purpose of the study was to investigate secular trends in cardiovascular disease risk factors over the twenty-year period 1960-1980 in the United States among young adult women, aged 18-24 years. This data previously appears in Basu and Landis (1995). In the study, the body mass index (BMI), expressed as weight (kg)/height $(m)^2$ has been considered as a risk factor of diastolic blood pressure (DBP). A DBP value exceeding 82.6 mmHg (determined from the 90th percentile of the distribution) is considered as hypertension. The 966 subjects are cross-classified with respect to the body mass index having 4 levels 0, 1, 2 and 3 (0: BMI<23, 1: $23 \le$ BMI< 25, 2: $25 \le$ BMI< 27 and 3: BMI ≥ 27) and the status of the hypertension (0: absence, 1: presence).

Table 2: Distribution of 966 subjects into four exposure levels with the respective disease status

BMI levels	Hypertension		Total
	0 (absence)	1 (presence)	
0: (BMI <23)	590	50	640
1: ($23 \le$ BMI< 25)	119	11	130
2: ($25 \le$ BMI< 27)	69	8	77
3: (BMI ≥ 27)	80	39	119
Total	858	108	966

For data set in Table 2, we calculate the MLEs p_{jk} of π_{jk} and based on these estimates we obtain the MLEs $\hat{AR}_j$, $j = 1, 2, 3$, and the corresponding 95% confidence intervals using Wald's test statistic. These results have been summarized in Table 3.

Table 3: The MLEs of overall and category-specific AR and the corresponding 95% confidence intervals

Parameter	Estimated value	Confidence interval
AR	0.3012	(0.1671, 0.4353)
AR_1	0.0078	(-0.0550, 0.0706)
AR_2	0.0184	(-0.0325, 0.0692)
AR_3	0.2750	(0.1784, 0.3716)

The estimate of overall attributable risk, AR, has been found by collapsing over all the exposure levels and constructing a 2×2 contingency table of exposure and disease factor. The estimated value of AR is 0.3012 and a 95% confidence interval is found to be (0.1671, 0.4353). Therefore, one can interpret that 30.1% of all hypertension can be attributed to BMI and could potentially be prevented by reducing BMI to less than 23. The attributable risk AR_1 for reducing level from 1 to level 0 is estimated at 0.0078 with a corresponding 95% confidence interval estimate of (-0.0550, 0.0706). This means that about 0.78% of all hypertension can be attributed to BMI and could potentially be

prevented by reducing BMI from level 1 ($23 \leq$ BMI< 25) to level 0 (BMI <23). The similar interpretations go for the estimates of AR_2 and AR_3 reported in Table 3.
Note that lower limit of the confidence intervals for both AR_1 and AR_2 is less than zero. This result suggests that there is no significant evidence at 5% to support that the proportional reduction in cardiovascular disease would be positive if the BMI reduced from level 1 to 0 ($23 \leq$ BMI< 25 to BMI <23), and from level 2 to 0 ($25 \leq$ BMI< 27 to BMI <23).

In order for further investigation regarding significance of a level, we are interested to estimate AR_{ij} considering intermediate level as the baseline level, and the corresponding confidence interval by using Wald's test statistic. The estimates of parameters AR_{12}, AR_{13} and AR_{23} are obtained to be 0.0137, 0.2679, and 0.2466, respectively which are reported with corresponding 95% confidence interval estimates in Table 4. From the estimate of AR_{12}, it can be interpreted that about 1 per cent of the risk of hypertension could be avoided by reducing the BMI from $25 \leq$ BMI< 27 to $23 \leq$ BMI< 25. Note that the lower limit for the confidence interval for AR_{12} is less than 0, which suggests that there is no significant evidence at 5% to support that the proportional reduction of the risk of hypertension would be positive if the BMI level reduces from level 2 to 1 based on this particular data. Therefore, one can combine levels 1 and 2 together to make a level $23 \leq$ BMI<27 and see if it is significant with respect to level 0. Likewise, using the point estimate of AR_{13}, it can be interpreted that about 27 per cent of the risk of hypertension could be avoided by reducing the BMI from BMI ≥ 27 to $23 \leq$ BMI<25. From the 95% confidence interval estimate of AR_{13}, it can be asserted that between 16 per cent and 38 per cent of the risk of developing hypertension could be eliminated by reducing the BMI from level 3 to 1. Similar interpretation goes for the estimate of AR_{23} and it's 95% confidence interval estimate reported in Table 4.

Table 4: The estimates of intermediate base-level attributable risk AR_{jk} and their 95% confidence intervals

Parameter	Estimated value	Confidence interval
AR_{12}	0.0137	(-0.0457, 0.0732)
AR_{13}	0.2679	(0.1601, 0.3757)
AR_{23}	0.2466	(0.1258, 0.3675)

4. Conclusion

The level-specific AR for a risk factor with multiple exposure levels may play an important role in targeting the population group with a higher AR while planning disease prevention strategies. For example, Coughlin et al. (1994) studied esophageal cancer for case-control data and found that AR for moderate alcohol drinkers (40-79 g/day) was 27%, whereas AR for heavy drinkers (120+ g/day) was lower (22%). This result suggests that the prevention strategies targeting for moderate drinkers would be more effective than those of heavy drinkers in that population. Also, while the categorization of a risk

factor is useful for epidemiologic study, it is also important to avoid unnecessary levels of an exposure. The study of AR with respect to intermediate level helps us determine the insignificant levels, and hence reducing the number of levels. The intermediate base-levels attributable risk also help in evaluating the proportion of the disease reduction between the intermediate levels, which is very important in disease prevention planning for a risk factor with a high exposure rate in the population of interest.

References

1. Agresti, A. (2002). *Categorical Data Analysis*. 2nd ed., Wiley, New York.
2. Basu, S. and Landis, J. R. (1995). Model-based estimation of population attributable risk under cross-sectional sampling. *Aamerican Journal of Epidemiology,* **142**: 1338-1343.
3. Cole, P. and MacMahon, B. (1971). Attributable risk percent in case-control studies. *British Journal of Preventive and Social Medicine,* **25**: 242-244.
4. Coughline, S. S., Benichou, J. and Weed, D. L. (1994). Attributable risk estimation in case-control studies. Epidemiologic Reviews, **16**: 51-64.
5. Denman, D. W. and Schlesselman, J. J. (1983). Interval estimation of the attributable risk for multiple exposure levels in case-control studies. *Biometrics*, **39**: 185-192.
6. Fleiss, J. L. (1979). Inference about population attributable risk from cross-sectional studies. *American Journal of Epidemiology*, **110**: 103-104.
7. Greenland, S. and Robins, J. M. (1988). Conceptual problems in the definition and interpretation of attributable fractions. *American Journal of Epidemiology*, **128**: 1185-1197.
8. Last, J. M. (1983). *A Dictionary of Epidemiology*. Oxford University Press, New York.
9. Levin, M. L. (1953). The occurrence of lung cancer in man. *Acta Unio Internationalis Contra Cancerum*, **9**: 531-541.
10. Lui, K. J. (2001). Confidence intervals of the attributable risk under cross-sectional sampling with confounders. *Biometrical Journal*, **43**: 767-779.
11. Lui, K. J. (2003). Interval estimation of the attributable risk for multiple exposure levels in case-control studies with confounders. *Statistics in Medicine*, **22**: 2443-2457.
12. McDowell, A., Engel, A., Massey, J. T. et al. (1981). Plan and Operation of the Second National Health and Nutrition Examination Survey, 1976-1980. Hyattsville, MD: National Center for Health Statistic. (Vital and health statistics, Ser. 1, no.15) (DHHS publication no. (PHS) 81-1317).
13. Miettinen, O. S. (1974). Proportion of disease caused or prevented by a given exposure, trait or intervention. *American Journal of Epidemiology*, **99**: 325-332.
14. Ouellet, B. L., Romeder, J. M. and Lance, J. M. (1979). Premature mortality attributable to smoking and hazardous drinking in Canada. *American Journal of Epidemiology*, **109**: 451-463.
15. Rao, C. R. (1973). *Linear Statistical Inference and Its Application*. 2nd ed., Wiley, New York.
16. Walter, S. D. (1976). The estimation and interpretation of attributable risk in health research. *Biometrics*, **32**: 829-849.