Noncollapsibility of common odds ratios without/with confounding

Eiji YAMAMOTO and Masayuki DOI

Okayama Univ. of Science, Okayama, Japan
Kannabe High School, Hiroshima, Japan

Introduction

Stratification analysis in contingency tables is used to control a potential confounding factor. Consider a follow-up study for an exposure E and a disease D, and let observations be stratified in I strata by a risk factor F. Within each stratum, suppose that the total number of individuals exposed and unexposed is n_i and m_i respectively, and the number of individuals who have the disease of exposed and unexposed is x_i and y_i, where $i = 1, \cdots, I$. Then, a $2 \times 2 \times I$ contingency table is constructed as follows:

F_i

	E	$\bar{E}$
D	x_i	y_i
$\bar{D}$	$n_i - x_i$	$m_i - y_i$
	n_i	m_i

,

	E	$\bar{E}$
D	$\sum x_i$	$\sum y_i$
$\bar{D}$	$\sum(n_i - x_i)$	$\sum(m_i - y_i)$
	$\sum n_i$	$\sum m_i$

If we ignore the risk factor F, the contingency tables (subtables) are collapsed to a 2×2 contingency table (crude table) as follows: Then, the stratum-specific odds ratios for the subtables and the crude odds ratio for the crude table can be estimated as

$$\widehat{OR}_i = \frac{x_i}{n_i - x_i} \Big/ \frac{y_i}{m_i - y_i} \qquad \text{and} \qquad \widehat{OR}_c = \frac{\sum x_i}{\sum(n_i - x_i)} \Big/ \frac{\sum y_i}{\sum(m_i - y_i)}$$

respectively.

Noncollapsibility

If the factor F is effective to D and the distribution of F in the exposed is different from that in the unexposed, we call it confounding factor or confounder. When the distribution of F in the exposed and in the unexposed are the same, the theorem 1 of the two strata case in Doi, Nakamura and Yamamoto (2001) can be extended in the general I stratum case that the odds ratio tends to the conservative value, that is, $OR_1 < OR_c < 1$ if $0 < OR_1 = \cdots = OR_I < 1$ or $1 < OR_c < OR_1$ if $OR_1 = \cdots = OR_I > 1$. Notice that there exists a difference between the common stratum-specific odds ratio and the crude odds ratio under no confounding, which is called noncollapsibility of the odds ratio.

Noncollapsibility and confounding are distinct concepts and must be distinguished clearly. Under the existence of the risk factor F, the difference between the odds ratios must be separated in two parts, the contribution of confounding and the contribution of noncollapsibility.

Noncollapsibility and confounding contribution of the difference between odds ratios

For $i = 1, \cdots, I$, let p_i and q_i be a proportion of diseases in the exposed and the unexposed within the stratum i respectively, and let $p(\alpha)$ and $q(\beta)$ be a proportion of diseases in the exposed and the unexposed in the overall population respectively, where $\alpha = (\alpha_1, \cdots, \alpha_I)$ and $\beta = (\beta_1, \cdots, \beta_I)$ are proportions of F_i in the exposed and unexposed respectively. Note that $p(\alpha) = \sum \alpha_i p_i$ and $q(\beta) = \sum \beta_i q_i$. These proportions can be estimated by

$$\hat{\alpha}_i = \frac{n_i}{\sum n_i},\ \hat{\beta}_i = \frac{m_i}{\sum m_i},\ \hat{p}_i = \frac{x_i}{n_i},\ \hat{q}_i = \frac{y_i}{m_i},\ \hat{p}(\hat{\alpha}) = \frac{\sum x_i}{\sum n_i},\ \hat{q}(\hat{\beta}) = \frac{\sum y_i}{\sum m_i}.$$

The stratum-specific odds ratio within the stratum $i = 1, \cdots, I$ and the crude odds ratio are defined as

$$OR_i = \frac{p_i/(1-p_i)}{q_i/(1-q_i)} \qquad \text{and} \qquad OR_c = \frac{p(\alpha)/(1-p(\alpha))}{q(\beta)/(1-q(\beta))}.$$

respectively, and may be estimated by $\widehat{OR}_i$ and $\widehat{OR}_c$ as mentioned previously. Here, it is assumed that the stratum-specific odds ratio is common, that is, $OR_1 = \cdots = OR_I$. The crude odds ratio depends on α and β and we can let $OR_c = OR_c(\alpha, \beta)$. When the distribution of F in the exposed and in the unexposed are the same, that is, the equation $\alpha = \beta$ holds, we will separate the difference between the crude odds ratio and the common odds ratio $OR_c(\alpha, \beta) - OR_1$ into two parts, the noncollapsibility contribution NC and the confounding contribution CC: $NC = OR_c(\alpha, \alpha) - OR_1$, $CC = OR_c(\alpha, \beta) - OR_c(\alpha, \alpha)$.

Confounding

From the definition of CC, it reduces to

$$CC = \frac{p(\alpha)}{1-p(\alpha)}\left(\frac{1}{q(\beta)} - \frac{1}{q(\alpha)}\right).$$

Thus, we may apply estimates $\hat{p}(\hat{\alpha})$, $\hat{q}(\hat{\alpha})$ and $\hat{q}(\hat{\beta})$ to appropriate points. As $q(\alpha) = \sum \alpha_i q_i$, the confounding contribution of the difference between odds ratios can be estimated as

$$\widehat{CC} = \frac{\sum x_i}{\sum(n_i - x_i)}\left(\frac{\sum m_i}{\sum y_i} - \frac{\sum n_i}{\sum n_i y_i/m_i}\right).$$

The sufficient condition of $\widehat{CC} = 0$, which indicates there is no confounding contribution in the bias of the crude odds ratio and the bias is due to noncollapsibility contribution only, is $n_i/m_i = \sum n_i / \sum m_i$ or $y_i/m_i = \sum y_i / \sum m_i$. These indicates that if the exposure E is independent of the factor F or F is not a risk factor among unexposed $\bar{E}$, the factor F can not be a confounder.

References

[1] Doi, M., Nakamura, T. and Yamamoto, E. (2001), Conservative tendency of the crude odds ratio, *J. Japan Statist. Soc.* **31-1**, 1-19.