NIASRA

NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

National Institute for Applied Statistics Research Australia

University of Wollongong

Working Paper

14-14

Mantel-Haenszel Estimators of a Common Local Odds Ratio for Multiple Response Data

Thomas Suesse and Ivy Liu

Mantel-Haenszel Estimators of a Common Local Odds Ratio for Multiple Response Data

Thomas Suesse · **Ivy Liu**

Abstract For a two–way contingency table with categorical variables, local odds ratios are commonly used to describe the relationships between the row and column variables. The ordinary case has mutually exclusive cell counts, i.e., each subject must fit into one and only one cell. However, in many surveys respondents may select more than one outcome category. We discuss the maximum likelihood and Mantel–Haenszel estimators of an assumed common local odds ratio for several $2 \times c$ tables, treating the multiple responses as an extension of the multinomial sampling model. We derive new dually consistent (co)variance estimators for the Mantel–Haenszel local odds ratio estimators and show their performance in a simulation study.

Keywords Consistency, Local odds ratio, Mantel–Haenszel estimator, Odds ratio, Multiple responses

1 Introduction

Many studies are designed to compare groups on a multi–level response variable. One often uses a two-way contingency table that cross–classifies subjects on both group and response variables to display relationships between them. A set of odds ratios, such as *local odds ratios* (Agresti 2013, p.54) that use four cells in adjacent rows and columns, can describe the associations. If a study attempts to control for other factors that might influence the relationships, a three-way contingency table can show the associations between the group and response variables controlling for a possibly confounding variable. When the confounding variable has many categories, such data are often sparse and the three-way table might consist of many small cell counts.

Thomas Suesse
School of Mathematics and Applied Statistics, University of Wollongong, Australia
E-mail: tsuesse@uow.edu.au

Ivy Liu
School of Mathematics, Statistics and Operations Research
Victoria University of Wellington, New Zealand

Furthermore, the multi–level response categories might not be mutually exclusive. For example, in a survey, respondents may select any number out of the outcome categories. The respondents are often told to "mark all that apply". The analysis of this type of data, called *multiple response* data, has received much attention since Loughin and Scherer (1998). This paper proposes a Mantel-Haenszel-type method to summarize the associations across strata and to provide statistical inferences on it for multiple response sparse data. We also compare different approaches for such data.

Let $\pi_{j|ik}$ be the probability of selecting item (or column) $j = 1, 2, \ldots, c$ for a subject in group (or row) $i = 1, 2$ and stratum $k = 1, \ldots, K$. We consider the following generalised linear models

$$\log(\pi_{j|ik}) = \alpha_{jk} + \beta_{ij} \tag{1}$$

and

$$\log(\pi_{j|ik}) = \gamma_{ik} + \alpha_{jk} + \beta_{ij} \tag{2}$$

The first implies that given items, rows and strata are independent. The second model has no three-factor interactions, that is, the association between any two remains constant across different levels of the third variable. Both models imply a common odds ratio across K strata, i.e. $\Psi_{jh} = \Psi_{jh|1} = \cdots = \Psi_{jh|K}$, where the kth odds ratio for items j and h is defined as

$$\Psi_{jh|k} = \frac{\pi_{j|1k}\pi_{h|2k}}{\pi_{j|2k}\pi_{h|1k}}. \tag{3}$$

The local odds ratio usually only refers to the particular setting: $h = j + 1$. Besides the odds ratio, the relative risk for stratum k and item j, $\theta_{j|k} := \pi_{j|1k}/\pi_{j|2k}$ is also popular to describe the association. Assume that the conditional association remain the same given the control variable, i.e. $\theta_j = \theta_{j|1} = \cdots = \theta_{j|K}$ or $\Psi_{jh|1} = \cdots = \Psi_{jh|K}$, the Mantel–Haenszel (MH) (1959) estimator is often used not only when the common relative risk/odds ratio assumption seems plausible, but also as a summary measure when the association varies only mildly across the tables.

For an ordinary case where the response categories are mutually exclusive, Greenland (1989) proposed the MH-type common local odds ratio and common relative risk estimators. They are *dually consistent*, i.e. consistent under the large–stratum (K is bounded while the number of subjects per stratum goes to infinity) and sparse–data (K goes to infinity with sample size, but the number of subjects per stratum remains fixed) limiting models. It is efficient under the null of no association.

For multiple response data, Loughin and Scherer (1998) proposed a weighted chi–square test and a bootstrap test for the hypothesis that the probability of selecting any given item is identical among levels of a predictor variable. A series of work by Decady and Thomas (2000), Bilder et al (2000), and Bilder and Loughin (2001; 2002) focused on tests of various hypotheses for a single multiple response variable. Later on, Thomas and Decady (2004) and Bilder and Loughin (2004) considered the tests of independence between two multiple response variables cases. Besides the tests, Agresti and Liu (1999; 2001) discussed different strategies for modeling multiple response data. Bilder and Loughin (2009) extended their earlier work to simultaneously model and estimate the association structure between two multiple

response variables in complex survey sampling situations. In addition to the modeling and testing procedures, Liu and Suesse (2008) derived a closed form of the odds ratio estimation by comparing the odds of each of the items being selected for different groups. None of the existing literature discusses the local odds ratio estimators (3) for sparse data.

The meaning of odds ratio Ψ_{jh} for multiple responses is slightly different from the ordinary case. Consider one stratum only. Let's express Ψ_{jh} as

$$\Psi_{jh} = \frac{\tilde{\pi}_{j|1}}{1 - \tilde{\pi}_{j|1}} \left(\frac{\tilde{\pi}_{j|2}}{1 - \tilde{\pi}_{j|2}} \right)^{-1}$$

with $\tilde{\pi}_{j|i} = \pi_{j|i}/(\pi_{j|i} + \pi_{h|i})$. Define Y_j to be the (yes=1, no=1) response on item j. When the sampling scheme is multinomial, then $\tilde{\pi}_{j|i}$ is the probability of $Y_j = 1$ conditional on $Y_j = 1$ or $Y_h = 1$ for a subject in row i. Usually, the term $\tilde{\pi}_{j|i}/(1 - \tilde{\pi}_{j|i})$ is described as the odds of choosing item j rather than item h. The odds ratio Ψ_{jh} is the ratio of two odds for row 1 against row 2.

For multiple responses, $\tilde{\pi}_{j|i}$ does not have the same meaning as above, because it is possible that $Y_j = Y_h = 1$. It happens when a subject selects both items j and h. Therefore, $\tilde{\pi}_{j|ik}$ becomes the proportion of $\pi_{j|i}$ relative to the sum of $\pi_{j|i}$ and $\pi_{h|i}$. We interpret $\tilde{\pi}_{j|i}/(1 - \tilde{\pi}_{j|i})$ as the odds of observing a positive response on item j rather than item h. The odds ratio Ψ_{jh} is the ratio of these two odds for row 1 against row 2. This interpretation is broader and it can be applied for the ordinary case as well. Alternatively, Ψ_{jh} has the meaning of a ratio of two relative risks:

$$\Psi_{jh} = \frac{\pi_{j|1}\pi_{h|2}}{\pi_{j|2}\pi_{h|1}} = \frac{\pi_{j|1}/\pi_{j|2}}{\pi_{h|1}/\pi_{h|2}} = \frac{\theta_j}{\theta_h},$$

for both ordinary and multiple response cases.

The aim of this paper is to explore efficient ways of estimating the common local odds ratios Ψ_{jh} and extend their inferences to describe the associations in a $2 \times c \times K$ table when we allow multiple responses for columns. As an example, more than 75 million surgical patients world-wise received anesthesia and if left untreated a large proportion will develop a combination of nausea and vomiting, after surgery (Carter et al 2009). In this situation nausea and vomiting are the two items/columns in a contingency table. To find the relationship between the type of anesthesia used (e.g., propofol or volatile anesthetic) during surgery and the symptom after surgery, the traditional statistical inference for contingency tables is invalid, because the numbers of subjects lie within these items are not mutually exclusive. Possible outcomes are no nausea/no vomiting, no nausea/vomiting, nausea/no vomiting and nausea/vomiting. The number of patients for each outcomes are denoted by X^{00}, X^{01}, X^{10} and X^{11}, where the superscript refers to a binary outcome (no, yes) for nausea and vomiting. Thus, the number of patients suffering nausea is $X_1 = X^{10} + X^{11}$ and the number of patients suffering vomiting is $X_2 = X^{01} + X^{11}$. The cell counts X_1 and X_2 are not mutually exclusive. For a $2 \times c \times K$ table, we use notation $X_{jh|ik}^{ab}$ to represent the number of subjects in row i and stratum k, responding a for item j and b for item h, where $a, b = \{0, 1\}$. Similarly, notations π^{00}, π^{01}, π^{10} and π^{11} denote the

corresponding probabilities. We assume that X^{00}, X^{01}, X^{10} and X^{11} are multinomially distributed. Carter et al (2009) proposed an estimator of the relative risk θ_j for item j to investigate which drug is more effective when aiming at minimising adverse effects. We focus on estimation of the conditional local odds ratio $\Psi_{jh|k}$ given confounding variables using the MH method.

Section 2 introduces two MH estimators. One is based on relative risk estimators and the other is based on the odds ratio estimation by taking the multiple response nature of the data into account. In Section 3, we discuss the model based methods, including the maximum likelihood (ML) and the generalised estimating equations. In Section 4, we illustrate methods using two examples. Section 5 shows the performance of our new estimators in a simulation study. The paper finishes with comments and discussions.

2 Mantel-Haenszel Estimators

2.1 Relative Risk Estimators

Model (1) implies a common relative risk, i.e. $\theta_j = \theta_{j|k}$ for all $j = 1, \ldots, c$ and $k = 1, \ldots, K$.

Let $X_{j|ik}$ be the number of positive responses for item j, row i and stratum k and let n_{ik} be the total number of subjects for row i and stratum k. A dually consistent MH estimator for estimating a common relative risk has been independently proposed by Nurminen (1981) and Kleinbaum et al (1982)

$$\hat{\theta}_j = \frac{C_{j|12}}{C_{j|21}}, \tag{4}$$

where $C_{j|ab} = \sum_{k=1}^K c_{j|abk}$ with $c_{j|abk} = X_{j|ak}d_{bk}$ and $d_{bk} = n_{bk}/N_k$. The notation n_{bk} is the bth row marginal total and N_k is the total sample size, for stratum k. For simplicity we only consider two rows, so $N_k = n_{1k} + n_{2k}$. In general, it would be defined as $N_k = \sum_a n_{ak}$.

A dually consistent variance estimator for $L_j := \log \theta_j$ was given by Greenland (1989)

$$\widehat{\mathrm{Var}(L_j)} = \frac{\sum_k c_{j|12k}d_{2k}}{2C_{j|12}^2} + \frac{\sum_k c_{j|12k}d_{1k} + c_{j|21k}d_{2k}}{2C_{j|12}C_{j|21}} + \frac{\sum_k c_{j|21k}d_{1k}}{2C_{j|21}^2}, \tag{5}$$

which estimates the asymptotic variance

$$\frac{\sum_k \frac{n_{1k}n_{2k}}{N_k^2}\left[(n_{1k}\pi_{j|1k}(1-\pi_{j|1k})) + (n_{2k}\pi_{j|2k}(1-\pi_{j|2k}))\theta_j^2\right]}{(\sum_k \frac{n_{1k}n_{2k}\pi_{j|2k}}{N_k})^2} \tag{6}$$

Under a common relative risk assumption $\theta_j = \theta_{j|k}$, a dually consistent MH estimator for Ψ_{jh} is

$$\hat{\Psi}_{jh}^* = \frac{\hat{\theta}_j}{\hat{\theta}_h} \tag{7}$$

The dual consistency follows because under model (1) the MH estimator $\hat{\theta}_h$ is dually consistent for θ_h. Surprisingly, it turns out that $\hat{\Psi}_{jh}^*$ is also dually consistent under Model (2), even though $\hat{\theta}_h$ is not consistent under the sparse-data limiting model. Appendix A shows that because $\hat{\Psi}_{jh}^*$ is a ratio of $\hat{\theta}_j$ and $\hat{\theta}_h$, both terms are biased with the same rate and the bias cancels out.

In order to find a dually consistent variance for $\hat{\Psi}_{jh}^*$, we note that

$$\mathrm{Var}\left(\log \hat{\Psi}_{jh}^*\right) = \mathrm{Var}(L_j) + \mathrm{Var}(L_h) - 2\mathrm{Cov}(L_j, L_h).$$

The estimators for the first two terms are given in (5). Greenland (1989) proposed an estimator $\widehat{\mathrm{Cov}(L_j, L_h)}$ when cell counts X's are mutually exclusive. For the multiple response data, we propose the estimator

$$\widehat{\mathrm{Cov}(L_j, L_h)} = \frac{\sum_{k=1}^K \left[\frac{n_{bk}^2}{N_k^2}(X_{jh|ak} - d_{jh|ak}) + \hat{\theta}_j\hat{\theta}_h \frac{n_{ak}^2}{N_k^2}(X_{jh|bk} - d_{jh|bk}) \right]}{C_{j|ab}C_{h|ab}} \quad (8)$$

The appendix B shows arguments for the dual consistency.

2.2 Odds Ratio estimators

Alternatively, we propose another dually consistent MH estimator of Ψ_{jh} using odds ratio estimators as follows:

$$\hat{\Psi}_{jh} = \frac{C_{jh}}{C_{hj}},$$

where $C_{jh} = \sum_{k=1}^K c_{jh|k}$ with $c_{jh|k} = X_{j|1k}X_{h|2k}/N_k$. See Appendix C for a proof of the dual consistency.

Next, we show that under the multiple response data, the dually consistent variance estimator has an additional term added to the estimator given by Greenland (1989). The additional term could be considered as the extra information for the multiple response added to the ordinary contingency table with mutually exclusive cell counts.

Let $L_{jh} = \log \hat{\Psi}_{jh}$. For the ordinary case, Greenland (1989) proposed the following variance estimator U_{jhh}^{old} for $\mathrm{Var}(L_{jh})$ and the following covariance estimator U_{jhs}^{old} for $\mathrm{Cov}(L_{jh}, L_{js})$:

$$U_{jhh}^{old} := \frac{\sum_k c_{jh|k}d_{jh|k}}{2C_{jh}^2} + \frac{\sum_k c_{hj|k}d_{hj|k}}{2C_{hj}^2} + \frac{\sum_k c_{jh|k}d_{hj|k} + c_{hj|k}d_{jh|k}}{2C_{jh}C_{hj}}$$

$$U_{jhs}^{old} := \frac{\sum_k X_{j|1k}X_{h|2k}X_{s|2k}/N_k^2}{3C_{jh}C_{js}} + \frac{\sum_k X_{j|+k}X_{h|2k}X_{s|1k}/N_k^2}{3C_{jh}C_{sj}}$$

$$+ \frac{\sum_k X_{j|+k}X_{h|1k}X_{s|2k}/N_k^2}{3C_{hj}C_{js}} + \frac{\sum_k X_{j|2k}X_{h|1k}X_{s|1k}/N_k^2}{3C_{hj}C_{sj}}$$

with $d_{jh|k} := (X_{j|1k} + X_{h|2k})/N_k$, and $X_{j|+k} = \sum_i X_{j|ik}$.

Under the multiple response case, we assume $\mathbf{X}_{jh|ik} = (X_{jh|ik}^{00}, X_{jh|ik}^{01}, X_{jh|ik}^{10}, X_{jh|ik}^{11})$ follows a multinomial distribution with parameters n_{ik} and $\boldsymbol{\pi}_{jh|ik} = (\pi_{jh|ik}^{00}, \pi_{jh|ik}^{01}, \pi_{jh|ik}^{10}, \pi_{jh|ik}^{11})$ with $\pi_{jh|ik}^{00} + \pi_{jh|ik}^{01} + \pi_{jh|ik}^{10} + \pi_{jh|ik}^{11} = 1$. The marginal probabilities can be computed from the pairwise probabilities by $\pi_{j|ik} = \pi_{jh|ik}^{10} + \pi_{jh|ik}^{11}$ and $\pi_{h|ik} = \pi_{jh|ik}^{01} + \pi_{jh|ik}^{11}$. We can now show that

$$\mathbb{E}X_{j|ik}X_{h|ik} = n_{ik}n_{ik}'\pi_{j|ik}\pi_{h|ik} + n_{ik}\pi_{jh|ik}^{11} \tag{9}$$

with $n_{ik}' = n_{ik} - 1$. If each subject can only choose one outcome category, following the multinomial samplings, we have $\text{Cov}(X_{j|ik}, X_{h|ik}) = -n_{ik}^2\pi_{j|ik}\pi_{h|ik}$ and $\mathbb{E}X_{j|ik}X_{h|ik} = n_{ik}n_{ik}'\pi_{j|ik}\pi_{h|ik}$. The ordinary (multinomial) case is a special case of multiple responses. In Appendix D we use these results to present a sketch of the proof that the new estimators U_{jhh} for $\text{Var}(L_{jh})$, U_{jhs} for $\text{Cov}(L_{jh}, L_{js})$ and U_{jhts} for $\text{Cov}(L_{jh}, L_{ts})$ are dually consistent for multiple response data. For convenience, denote $X_{jh|ik}^{11}$ by $X_{jh|ik}$. The estimators U_{jhh}, U_{jhs} are defined as follows:

$$U_{jhh} := \widehat{\text{Var}}(L_{jh}) = U_{jhh}^{old} + U_{jhh}^{add}$$
$$U_{jhs} := \widehat{\text{Cov}}(L_{jh}, L_{js}) = U_{jhs}^{old} + U_{jhs}^{add}, \tag{10}$$

where the additional terms U^{add} are given by

$$U_{jhh}^{add} = -4\frac{\sum_k X_{j|1k}X_{h|1k}X_{jh|2k}/N_k^2 + \sum_k X_{jh|1k}X_{j|2k}X_{h|2k}/N_k^2}{2C_{jh}C_{hj}}$$
$$- \frac{\sum_k\{X_{jh|1k}(X_{j|2k} + X_{h|2k}) + X_{jh|2k}(X_{j|1k} + X_{h|1k})\}/N_k^2}{2C_{jh}C_{hj}}$$
$$+ 4\frac{\sum_k X_{jh|2k}X_{jh|1k}/N_k^2}{2C_{jh}C_{hj}}$$

and

$$U_{jhs}^{add} = \frac{\hat{V}_{jhs|12}^A}{C_{jh}C_{js}} - \frac{\hat{V}_{jh,js}}{C_{hj}C_{js}} - \frac{\hat{V}_{js,jh}}{C_{jh}C_{sj}} + \frac{\hat{V}_{jhs|21}^A}{C_{hj}C_{sj}}$$
$$+ \frac{\hat{V}_{jhs|12}^B}{3C_{jh}C_{js}} + \frac{\hat{V}_{hjs|12}^B + \hat{V}_{sjh|21}^B}{3C_{hj}C_{js}} + \frac{\hat{V}_{sjh|12}^B + \hat{V}_{hjs|21}^B}{3C_{jh}C_{sj}} + \frac{\hat{V}_{jhs|21}^B}{3C_{hj}C_{sj}}$$

with

$$\hat{v}_{jhs|ijk}^A = \frac{1}{N_k^2}X_{j|ik}^2 X_{hs|jk}, \qquad \hat{v}_{jhs|ijk}^B = -\frac{1}{N_k^2}X_{j|ik}X_{hs|jk}$$

$$\hat{v}_{jh,ts|k} = \frac{1}{N_k^2}\{X_{j|1k}X_{h|1k}X_{ts|2k} + X_{jh|1k}X_{t|2k}X_{s|2k} - X_{jh|1k}X_{ts|2k}\}$$

and $\hat{V}$ representing $\sum_k \hat{v}_k$. The estimator $U_{jhts} := \widehat{\text{Cov}}(L_{jh}, L_{ts})$ is given by

$$U_{jhts} := \frac{\hat{V}_{jt,hs}}{C_{jh}C_{ts}} - \frac{\hat{V}_{ht,js}}{C_{hj}C_{ts}} - \frac{\hat{V}_{js,ht}}{C_{jh}C_{st}} + \frac{\hat{V}_{hs,jt}}{C_{hj}C_{st}}.$$

When each subject can only choose one outcome category, the pairwise observations $X_{jh|ik}$ are all zero, because it is impossible to have both items chosen. Consequently, $U_{jhh}^{add} = U_{jhs}^{add} = U_{jhts} = 0$, such that $U_{jhh} \equiv U_{jhh}^{old}$ and $U_{jhs} \equiv U_{jhs}^{old}$. This shows that our estimators are generalizations of Greenland's estimators and are also applicable for the multinomial sampling model in an ordinary case with only one response outcome for each subject.

3 Model-based Estimators

Agresti and Liu (1999; 2001) discussed model-based strategies for multiple response data. This paper investigates two estimation strategies – the generalised estimating equations (GEE) (Liang and Zeger 1986) and the maximum likelihood (ML). We compare these two model-based estimators with the proposed MH estimators.

3.1 GEE method

Models (1) and (2) have the form of a generalized linear model (GLM) (McCullagh and Nelder 1989), however observations are not independent due to the nature of multiple response data. One method to deal with correlated data uses generalised estimating equations (GEE) (Liang and Zeger 1986). The user can choose between some common options (e.g. independence, exchangeable and unstructured) of the so-called working correlation. Independently of the working correlation, the parameter estimates of the mean model are still consistent.

3.2 ML method

When $c = 2$, assume that the complete data $\mathbf{X}_{jh|ik} = \{X_{jh|ik}^{00}, X_{jh|ik}^{01}, X_{jh|ik}^{10}, X_{jh|ik}^{11}\}$ in each row and stratum follow a multinomial distribution, with multinomial cell probabilities $\{\pi_{jh|ik}^{00}, \pi_{jh|ik}^{01}, \pi_{jh|ik}^{10}, \pi_{jh|ik}^{11}\}$. For $c > 2$, the complete data are formed by the counts of 2^c possible response sequences, according to the (no, yes) response for each item category. The maximum likelihood (ML) theory for Models (1) and (2) requires that the multinomial likelihood based on the complete data is maximized under the constraints imposed by the model for the marginal probabilities $\{\pi_{j|ik}, i = 1, 2, j = 1, \ldots, c, k = 1, \ldots, K\}$.

Haber (1985) and Lang and Agresti (1994) presented numerical algorithms for maximizing multinomial likelihoods subject to constraints for generalized loglinear models having the matrix form

$$\mathbf{C} \log \mathbf{A}\mathbf{p} = \mathbf{Z}\boldsymbol{\beta}, \tag{11}$$

where $\mathbf{p}$ refers to the vector of all multinomial cell probabilities, such as $\pi_{jh|ik}^{00}, \pi_{jh|ik}^{01}, \pi_{jh|ik}^{10}, \pi_{jh|ik}^{11}$. Under the assumption of a common odds ratio, the matrix $\mathbf{A}$ contains 0 and 1 entries in such a pattern that when applied to $\mathbf{p}$ it forms the relevant marginal

probabilities $\pi_{j|ik}$; the matrix $\mathbf{C}$ contains 0, 1, and -1 entries, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{c-1})'$ and $\mathbf{Z}$ is a row vector of K 1's.

An R function (mph.Rcode.R) for the algorithm may be obtained from Prof J. B. Lang of the Statistics Department, University of Iowa http://www.stat. uiowa.edu/~jblang/. Bergsma et al (2009) proposed another fitting algorithm available in R (called "cmm"), a modification of the Lang–Agresti algorithm (Lang and Agresti 1994, Lang 1996). Our simulation study used the R package "cmm" for the ML approach.

4 Simulation Study

This section investigates the performance of common local odds ratio estimators for MH, GEE, and ML approaches under Models (1) and (2). We simulated stratified multiple response data under Models (1) for which the common relative risk assumption holds and (2) under which the common relative risk assumption does not hold.

Let $c = 2$. For a fixed odds ratio $\Psi = \Psi_{12} = 1, 4$ and a fixed common relative risk $\theta_2 = 0.2$, and a fixed $\beta_{11} = 0.2$, the remaining β-coefficients are found from $\Psi = \frac{\beta_{11}\beta_{22}}{\beta_{12}\beta_{21}}$. The α_{jk} were simulated from a normal distribution with mean -1 and variance 0.3. This set-up applies under Models (1) and (2). Model (2) has the additional $\gamma_{ik} \sim N(-1, \sqrt{0.2})$. The α_{jk} and γ_{ik} are constrained by the condition that probabilities are bounded by 1. This provides the marginal probabilities $\pi_{h|ak}$ for $h = 1, 2$, $a = 1, 2$ and $k = 1, \ldots, K$. To measure the the pairwise dependency between items j and h we use the pair-wise odds ratio $\Gamma_{jh|ik}$, following Bilder and Loughin (2002):

$$\Gamma_{jh|ik} = \frac{P(Y_j = 1, Y_h = 1|ik)P(Y_j = 0, Y_h = 0|ik)}{P(Y_j = 0, Y_h = 1|ik)P(Y_j = 1, Y_h = 0|ik)}.$$

From the marginal probabilities $\{\pi_{j|ik}, j = 1, \ldots, c\}$ and the odds ratios $\{\Gamma_{jh|ik}, j \neq h = 1, \ldots, c\}$, we can compute the unique set of pairwise probabilities $\{\boldsymbol{\pi}_{jh|ik}, j \neq h = 1, \ldots, c\}$. Since we only consider $c = 2$ items, the pair-wise probabilities specify the joint distribution for given stratum and row.

Data were generated under $N_1 = \cdots = N_K$. Let K, N_k, Γ, and Ψ vary as $K = 1, 5, 20, 50, 100$, $N_k = 10, 20, 50, 100, 500$, $\Gamma_{12|ik} = 0.01, 1, 10$, and $\Psi = 1, 4$. For most settings $10,000$ data sets have been generated, except under $N_k = 50, 100$ for which $5,000$ data sets were generated to reduce the increased time needed for these settings.

Figures 1-4 compare the performance for the 4 different estimators including $\hat{\Psi}^*$ (Section 2.1), $\hat{\Psi}$ (Section 2.2), $\hat{\Psi}$ using the GEE approach (Section 3.1) , and $\hat{\Psi}$ using the ML approach (Section 3.2) under different scenarios. Figure 1 shows the relative mean squared error (mse) relative to the best estimator with the smallest mse. The value of 1 indicates the best method. Figure 4 gives the proportion of times that a 95% confidence interval of $\hat{\Psi}$ covers the true Ψ over 10,000 simulated data. For the MH estimators ($\hat{\Psi}^*$ and $\hat{\Psi}$), we present two types of confidence intervals. One is based on the derived dually consistent variance estimator and the other is based on the bootstrap-t method (Davison and Hinkley 1997), denoted by BT. Figure 4

shows the relative expected length relative to the shortest length for 95% confidence intervals. The results are shown under Model (1) and under Model (2). What is not shown is the non-convergence rate. In general, the ML and GEE methods often show a high rate of non-convergence. In contrast, the MH methods have a much lower non-convergence rate, and in particular $\hat{\Psi}^*$ has the lowest non-convergence rate. Notice that all confidence intervals and mse's were calculated in the logarithm scale, i.e., $ln(\hat{\Psi})$.

It is expected that the performance of the GEE and ML estimators is hardly better than those of the MH estimators, except for large N_k (large stratum situation) and possibly for the special case of $\Psi = 1$. For an ordinary three-way contingency table, Liu and Agresti (1996), Liu (2003) discussed that the ML parameter estimator for the conditional association between rows and columns given strata is not consistent under the sparse–data limiting model and the performance of the GEE estimator is not good as well if data are sparse (Liu and Suesse 2008). Surprisingly $\hat{\Psi}^*$ performs always slightly better than $\hat{\Psi}$ in terms of the mse. However $\hat{\Psi}$ is superior in the sense that the coverage is closer to the 95% compared to $\hat{\Psi}$ at a 5% significance level. This is expected because the dual consistency of the variance estimator of $\hat{\Psi}^*$ only applies under the common relative risk assumption (Model (1)). For both MH estimators $\hat{\Psi}^*$ and $\hat{\Psi}$, the bootstrap confidence interval is only better than the confidence interval based on the proposed variance estimators under Model (1) and under the large-stratum cases. Under the sparse-data situation and under Model (2), the newly proposed estimator $\hat{\Psi}$ along with variance estimators is preferred.

The common local odds ratio holds not only for Models (1) and (2), but also for the following model:

$$\log(\pi_{j|ik}) = \gamma_{ik} + \beta_{ij}$$

Without the loss of generality, we expect that the performance of GEE, ML and MH remains similar to Models (1) and (2).

5 Examples

This section illustrates our proposed estimators using two examples. The first example of combination of nausea and vomiting after surgery has only 1 stratum, see Table 2 in Carter et al (2009). The authors reported a relative risk for vomitting of 0.82 of propofol relative to volatile anesthetic. The relative risk for nausea is 0.81 of propofol relative to volatile anesthetic. A research question is whether there is a difference in the relative risk.

The estimates of the local odds ratio (or the ratio of relative risks), Ψ, and the corresponding 95% confidence intervals (shown in parenthesis) are $\hat{\Psi}^* = 1.0123$ $(0.9011, 1.1372)$, $\hat{\Psi} = 1.0123\,(0.8858, 1.1568)$, $\hat{\Psi}$-GEE $= 1.0123\,(0.8857, 1.1569)$, and $\hat{\Psi}$-ML $= 1.0096\,(0.9339, 1.0914)$. Carter et al (2009) gave the same results for the GEE method. Although these estimates are slightly different, they reach the same conclusion that there is no significant difference between propofol and volatile anesthetic on the two symptoms.
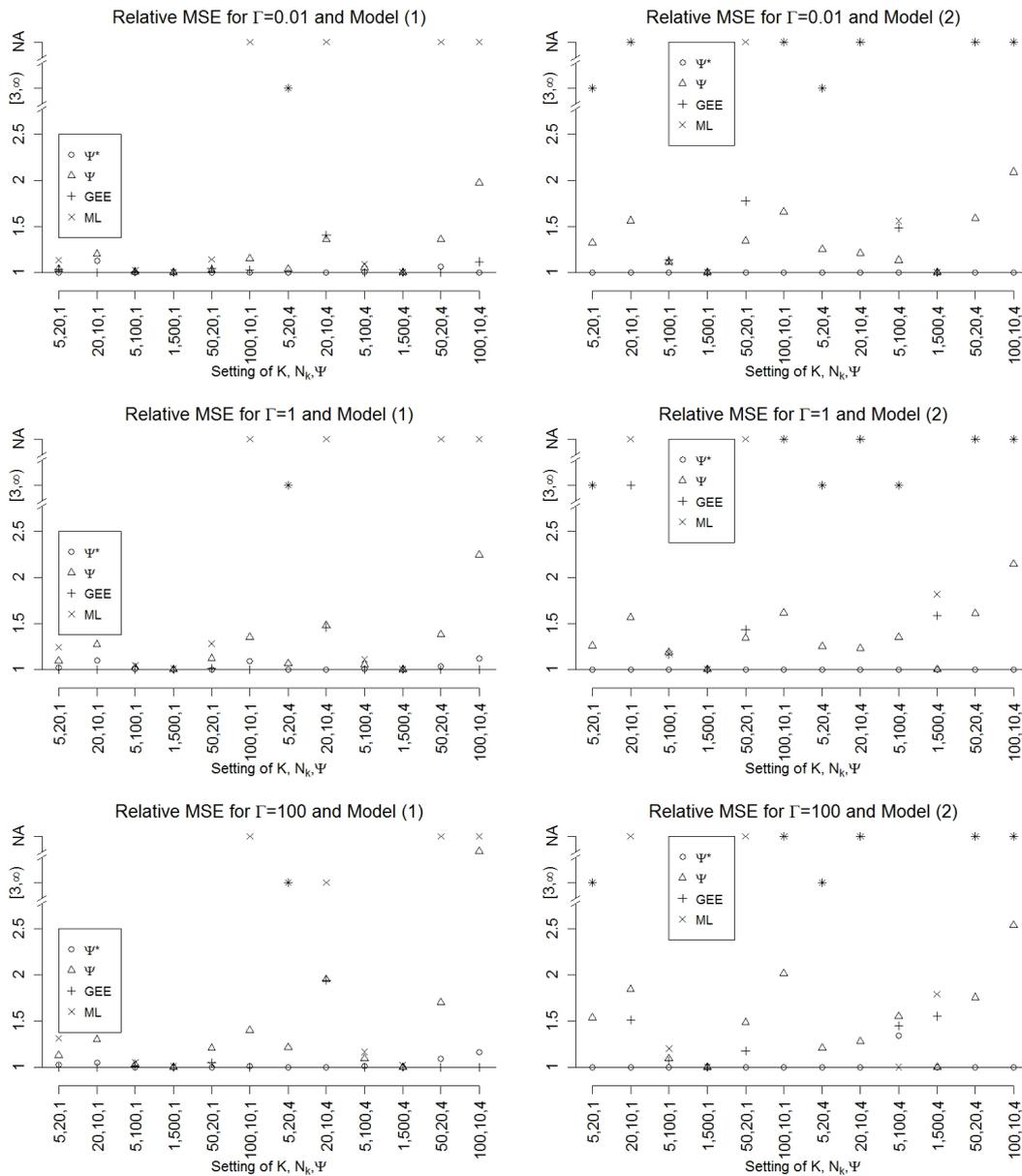
Fig. 1 Relative MSE for $\Gamma = 0.01, 1, 100$ for Model (1) (left) and Model (2) (right), 'NA' (not available) indicates that the estimator could never be calculated

The second example contains a highly stratified dataset, conducted by Gu et al (2005). The researchers are interested in the gender difference among English learning strategies used for primary school students in Singapore. This paper investigates
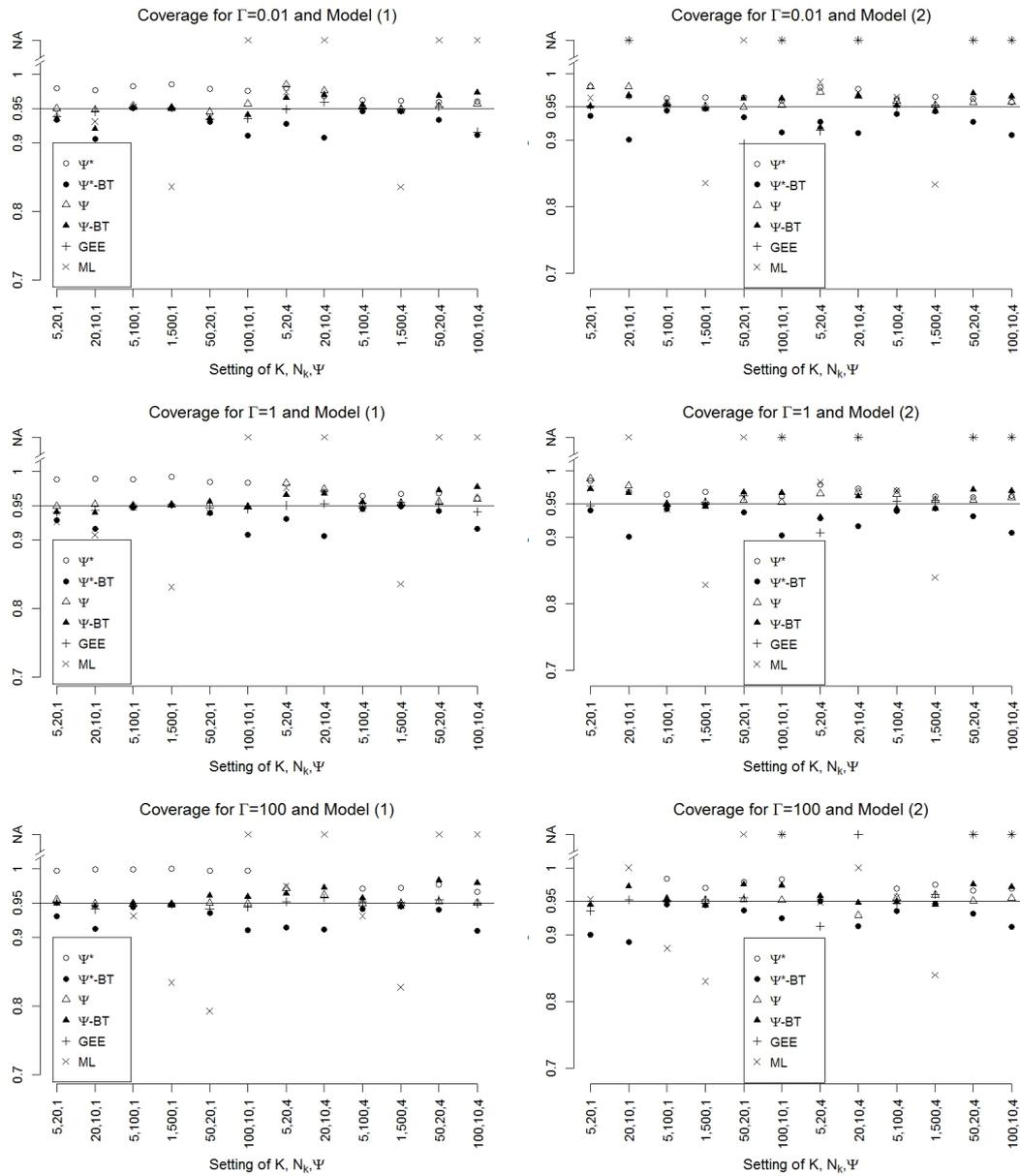
Fig. 2 Coverage for $\Gamma = 0.01, 1, 100$ for Model (1) (left) and Model (2) (right), 'NA' (not available) indicates that the estimator could never be calculated

the difference between girls and boys on five strategy questions related to English listening – Q1: "When I am free, I find interesting things to listen to in English (for example, TV, radio, etc)"; Q2: "After I finish listening, I make a summary in my
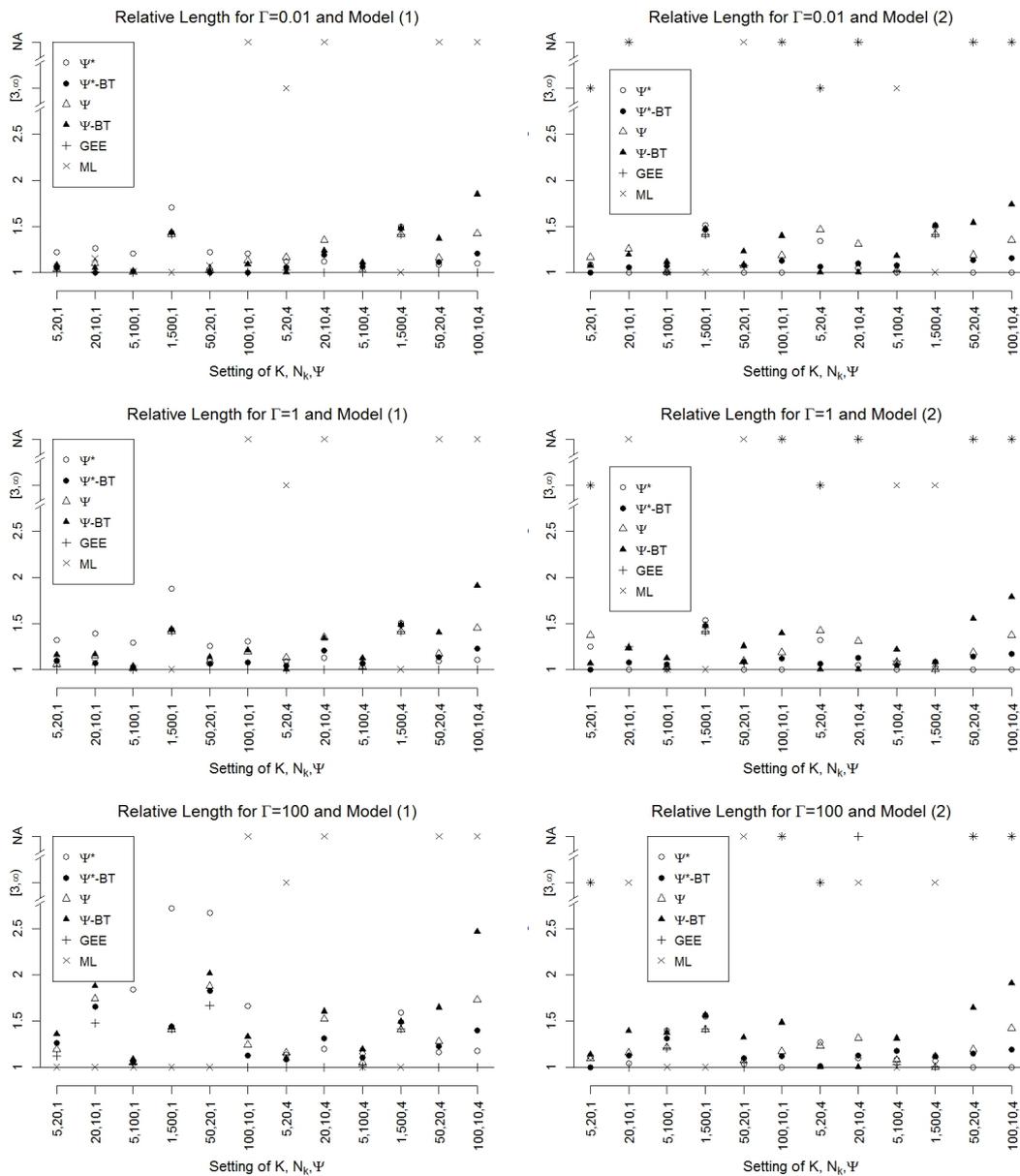
Fig. 3 Relative Length for $\Gamma = 0.01, 1, 100$ for Model (1) (left) and Model (2) (right), 'NA' (not available) indicates that the estimator could never be calculated

mind about what I heard"; Q3: "I tell myself to enjoy listening in English"; Q4: "When I don't understand something, I use my knowledge about the topic to guess"; and Q5: " When I listen, I repeat the pronunciation of the words I have heard". It

Gender (i)	Q1	Q2	Q3	Q4	Q5	n_i
boys (1)	246	198	253	209	191	537
girls (2)	298	217	266	170	221	530

Table 1 Marginal counts of positive responses for the fives questions at hand for boys and girls summarised over all $K = 150$ strata

	Q1	Q2	Q3	Q4	Q5
Q1	$--$	$-0.036(0.139)$	$-0.132(0.132)$	$-0.387(0.147)$	$-0.003(0.142)$
Q2	$-0.021(0.108)$	$--$	$-0.096(0.139)$	$-0.351(0.152)$	$0.033(0.147)$
Q3	$-0.130(0.087)$	$-0.004(0.107)$	$--$	$-0.254(0.150)$	$0.129(0.141)$
Q4	$-0.312(0.116)$	$-0.416(0.120)$	$-0.252(0.118)$	$--$	$0.384(0.155)$
Q5	$0.020(0.104)$	$0.012(0.118)$	$0.104(0.105)$	$0.281(0.13)$	$--$

Table 2 Estimates of local log-odds ratios $\hat{\Psi}^*$ (upper right) and $\hat{\Psi}$ (lower left) followed by standard errors in brackets

	Q1	Q2	Q3	Q4	Q5
Q1	$--$	$(0.693, 1.236)$	$(0.617, 1.135)$	$(0.390, 0.968)$	$(0.719, 1.275)$
Q2	$(0.767, 1.191)$	$--$	$(0.636, 1.18)$	$(0.406, 1.003)$	$(0.746, 1.321)$
Q3	$(0.706, 1.049)$	$(0.787, 1.205)$	$--$	$(0.482, 1.069)$	$(0.862, 1.414)$
Q4	$(0.505, 0.959)$	$(0.423, 0.896)$	$(0.545, 1.01)$	$--$	$(1.164, 1.771)$
Q5	$(0.817, 1.224)$	$(0.78, 1.244)$	$(0.904, 1.315)$	$(1.069, 1.579)$	$--$

Table 3 95% CI based on $\hat{\Psi}^*$ (upper right) and $\hat{\Psi}$ (lower left)

is well known that there are some differences between girls and boys on using English listening strategies (Gu 2002). The researchers are more interested to find out whether the gender difference varies across different strategies. We illustrate our proposed method to describe the relationship between gender and the five questions, by controlling on many possible confounding variables, such as school, ethnicity, and English level. Each stratified 2×5 table contains genders (Female and Male) in rows and the questions (Q1 – Q5) in columns. In total, there are 150 stratified tables.

The pairwise data $X_{jh|ik}^{00}, X_{jh|ik}^{01}, X_{jh|ik}^{10}, X_{jh|ik}^{11}$ represent the numbers of students whose response on (Q_j, Q_h) are (no, no), (no, yes), (yes, no), and (yes, yes), respectively for row i and stratum k. The cell counts in each 2×5 table are the marginal totals of $\mathbf{X}_{jh|ik}$ according to the yes and no responses on each question. For example, the cell count in row 1 and column 1 is $X_{12|1k}^{10} + X_{12|1k}^{11}$, and is $X_{12|1k}^{01} + X_{12|1k}^{11}$ in row 1 and column 2. Table 1 shows the marginal counts by collapsing over all 150 strata.

Both GEE and ML algorithms did not converge and cannot be used to obtain estimates of Ψ. The MH estimates for $\log \Psi_{jh}$ are given in Table 2 and Table 3 shows the 95% confidence intervals for Ψ_{jh}. The confidence intervals indicate that the gender difference is similar in Questions 1, 2, 3, and 5. Question 4 is significantly different from the rest of the questions based on both MH estimates.

6 Discussion

The article considers estimators of the local odds ratio under an extension of the ordinary case to the multiple response case. In the ordinary case, each of K $2 \times c$ contingency tables assumes two independent rows of multinomial samples and the response categories are mutually exclusive. In the multiple response case, the response categories are not necessarily mutually exclusive. For example in surveys, it is very common to tick all that apply and not only one item that applies.

Greenland (1989) proposed the MH estimators and their (co)variance estimators to summarize the conditional association between rows and columns in a 3-way table under the ordinary case. We discuss GEE, ML and MH methods applied to stratified multiple response data. This paper gives two types of MH estimators. One is derived from relative risk estimators and the other uses the form of the ordinary MH odds ratio estimator. Both of them perform well compared to the GEE and ML estimators, especially when data are sparse and under the general case $\Psi \neq 1$. These estimators provide effective alternatives in particular when the GEE and ML methods cannot be calculated, as the latter methods have a high rate of non-convergence.

Suesse and Liu (2012) showed that the odds ratio estimation for K $2 \times c$ tables based on c dependent binomials is an extension of the independent binomial sampling model presented by Greenland (1989). This paper further generalizes the MH (co)variance estimators to the multiple response situation showing that the Greenland (1989) (co)variance estimator is a special case of our newly proposed estimator.

The odds ratio has the following property $\Psi_{jh} = \Psi_{js} \Psi_{sh}$. Thus $\log \Psi_{jh}$ cannot only be estimated by L_{js} but also by $L_{js} + L_{sh}$. There is no unique estimator. Greenland (1989) proposed the following generalized MH estimator following the Mickey and Elashoff (1985) approach:

$$\widehat{\log \Psi_{jh}} := \bar{L}_{jh} := (L_{j+} - L_{h+})/c.$$

This approach is independent of the applied estimator and generally applicable to any estimator of $\log \Psi_{jh}$. Then, the generalized MH estimators $\{\bar{L}_{jh}\}$ have the property $\bar{L}_{jh} = \bar{L}_{js} + \bar{L}_{sh}$. Liu and Suesse (2008) derived variance and covariance estimators for the generalized MH-type global odds ratio estimators, as a function of U's. Replacing their U's with the newly proposed U's, we can obtain the variance and covariance estimators for the proposed MH estimator. The results shown for the examples in Section 4 are based on the generalized MH estimators.

This paper only considered K $2 \times c$ tables and could be further extended to K $r \times c$ (with $r > 2$) tables. This extension would lead to another generalized MH estimator and different formulae for the (co)variance estimators of these generalized MH estimators. These formulae also require additional unknown covariance estimators, which are subject to future research.

References

Agresti A (2013) Categorical Data Analysis, 3rd edn. Wiley Series in Probability and Statistics, Wiley, New Jersey

Agresti A, Liu I (2001) Strategies for modeling a categorical variable allowing multiple category choices. Sociological Methods & Research 29(4):403–434

Agresti A, Liu IM (1999) Modeling a categorical variable allowing arbitrarily many category choices. Biometrics 55(3):936–943

Bergsma W, Croon M, Hagenaars J (2009) Marginal Models for Dependent, Clustered, and Longitudinal Categorical Data. Springer, New York

Bilder CR, Loughin TM (2001) On the first-order rao-scott correction of the umesh-loughin-scherer statistic. Biometrics 57(4):1253–1255

Bilder CR, Loughin TM (2002) Testing for conditional multiple marginal independence. Biometrics 58(1):200–208

Bilder CR, Loughin TM (2004) Testing for marginal independence between two categorical variables with multiple responses. Biometrics 60(1):241–248

Bilder CR, Loughin TM (2009) Modeling multiple-response categorical data from complex surveys. Canadian Journal of Statistics-Revue Canadienne De Statistique 37(4):553–570

Bilder CR, Loughin TM, Nettleton D (2000) Multiple marginal independence testing for pick any/c variables. Communications in Statistics-Simulation and Computation 29(4):1285–1316

Carter R, Zhang X, Woolson R, Apfel C (2009) Statistical analysis of correlated relative risks. Journal of Data Science 7:397–407

Davison A, Hinkley D (1997) Bootstrap methods and their application. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Oxford

Decady YJ, Thomas DR (2000) A simple test of association for contingency tables with multiple column responses. Biometrics 56(3):893–896

Greenland S (1989) Generalized mantel-haenszel estimators for k 2xj tables. Biometrics 45(1):183–191

Gu PY, Hu G, Zhang LJ (2005) Investigating language learner strategies among lower primary school pupils in singapore. Language and Education 19(4):281–303

Gu Y (2002) Gender, academic major, and vocabulary learning strategies of chinese efl learners. RELC Journal 33(1):35–54

Haber M (1985) Maximum-likelihood methods for linear and log-linear models in categorical-data. Computational Statistics & Data Analysis 3(1):1–10

Kleinbaum D, Kupper L, Morgenstern H (1982) Epidemiologic Research: Principles and Quantitative Methods. Lifetime Learning Publications, Belmont, California

Lang JB (1996) Maximum likelihood methods for a generalized class of log-linear models. Annals of Statistics 24(2):726–752

Lang JB, Agresti A (1994) Simultaneously modelling joint and marginal distributions of multivariate categorical responses. Journal of the American Statistical Association 89(426):625–632

Liang KY, Zeger SL (1986) Longitudinal data-analysis using generalized linear-models. Biometrika 73(1):13–22

Liu I (2003) Describing ordinal odds ratios for stratified r x c tables. Biometrical Journal 45(6):730–750

Liu I, Suesse T (2008) The analysis of stratified multiple responses. Biometrical Journal 50(1):135–149

Liu IM, Agresti A (1996) Mantel-haenszel-type inference for cumulative odds ratios with a stratified ordinal response. Biometrics 52(4):1223–1234

Loughin TM, Scherer P (1998) Testing for association in contingency tables with multiple column responses. Biometrics 54:630–637

Mantel N, Haenszel W (1959) Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute 22(4):719–748

McCullagh P, Nelder JA (1989) Generalized Linear Models, 2nd edn. Chapman and Hall, New York

Mickey RM, Elashoff RM (1985) A generalization of the mantel-haenszel estimator of partial association for 2xjxk-tables. Biometrics 41(3):623–635

Nurminen M (1981) Asymptotic efficiency of general non-iterative estimators of common relative risk. Biometrika 68(2):525–530

Sen PK, Singer JM (1993) Large Sample Methods in Statistics: An Introduction with Applications. Chapman & Hall, New York

Suesse T, Liu I (2012) Mantelhaenszel estimators of odds ratios for stratified dependent binomial data. Computational Statistics and Data Analysis 56:2705–2717

Thomas DR, Decady YJ (2004) Testing for association using multiple response survey data: Approximate procedures based on the rao-scott approach. International Journal of Testing 4(1):43–59

A Dually Consistency of $\hat{\Psi}_{jh}^{*}$

$$\hat{\theta}_h = \frac{\sum_k X_{h|ak} \frac{n_{bk}}{N_k}}{\sum_k X_{h|bk} \frac{n_{ak}}{N_k}}$$

which converges to

$$\frac{\sum_k \pi_{h|ak} \frac{n_{bk}n_{ak}}{N_k}}{\sum_k \pi_{h|bk} \frac{n_{ak}n_{bk}}{N_k}} = \frac{\sum_k \exp(\gamma_{ak} + \alpha_{hk} + \beta_{ah}) \frac{n_{bk}n_{ak}}{N_k}}{\sum_k \exp(\gamma_{bk} + \alpha_{hk} + \beta_{bh}) \frac{n_{ak}n_{bk}}{N_k}}$$

$$= \frac{\exp(\beta_{ah}) \sum_k \exp(\gamma_{ak} + \alpha_{hk}) \frac{n_{bk}n_{ak}}{N_k}}{\exp(\beta_{bh}) \sum_k \exp(\gamma_{bk} + \alpha_{hk}) \frac{n_{ak}n_{bk}}{N_k}}$$

$$= \exp(\beta_{ah} - \beta_{bh}) \frac{\sum_k \exp(\gamma_{ak} + \alpha_{hk}) \frac{n_{bk}n_{ak}}{N_k}}{\sum_k \exp(\gamma_{bk} + \alpha_{hk}) \frac{n_{ak}n_{bk}}{N_k}}$$

Under model (1), the term on the right hand side becomes 1, and dual consistency applies because $\theta_h = \exp(\beta_{ah} - \beta_{bh})$. However under model (2), this is not the case because of the γ_{ik} terms. This means that the estimator $\hat{\theta}_h$ converges to $\theta_h \times c$, where c is a constant.

Now $\hat{\theta}_j$ converges under model (2) to $\theta_j \times c$, therefore $\hat{\Psi}_{jh}^{*} = \hat{\theta}_j / \hat{\theta}_h$ converges to $\frac{\theta_j \times c}{\theta_h \times c} = \Psi_{jh}$, and dual consistency even applies under model (2).

B Dually Consistency of Covariance Estimator of two MH relative risk Estimators

Showing that $\text{Cov}(L_j, L_h)$ is consistent is equivalent to showing that $\text{Cov}(\hat{\theta}_j, \hat{\theta}_h)$ is consistent by application of delta method to log-function. Hence we need to show that

$$\widehat{\text{Cov}(\hat{\theta}_j, \hat{\theta}_h)} = \frac{\frac{n_{bk}^2}{N_k^2}(X_{jh|ak} - d_{jh|ak}) + \hat{\theta}_j \hat{\theta}_h \frac{n_{ak}^2}{N_k^2}(X_{jh|bk} - d_{jh|bk})}{C_{j|ba} C_{h|ba}}$$

is consistent for $\text{Cov}(\hat{\theta}_j, \hat{\theta}_h)$.

We can show that

$$\lim \text{Cov}(\hat{\theta}_j - \theta_j, \hat{\theta}_h - \theta_h) = \lim \text{Cov}\left(\frac{C_{j|ab} - \theta_j C_{j|ba}}{C_{j|ba}}, \frac{C_{h|ab} - \theta_h C_{h|ba}}{C_{h|ba}}\right)$$

$$= \frac{\sum_k \text{Cov}(c_{j|ab} - \theta_j c_{j|ba}, (c_{h|ab} - \theta_h c_{h|ba}))}{\lim C_{j|ba} \lim C_{h|ba}}$$

and

$$\text{Cov}(c_{j|ab} - \theta_j c_{j|ba}, c_{h|ab} - \theta_h c_{h|ba}) = \frac{n_{ak} n_{bk}}{N_k^2}\left(n_{bk}(\pi_{jh|ak} - \pi_{j|ak}\pi_{h|ak}) + n_{ak}\theta_j \theta_h(\pi_{jh|bk} - \pi_{j|bk}\pi_{h|bk})\right)$$

which can be estimated under both limiting models by

$$\frac{n_{bk}^2}{N_k^2}(X_{jh|ak} - d_{jh|ak}) + \hat{\theta}_j \hat{\theta}_h \frac{n_{ak}^2}{N_k^2}(X_{jh|bk} - d_{jh|bk})$$

with $d_{jh|ak} = (X_{j|ak} X_{h|ak} - X_{jh|ak})/n'_{ak}$ and $n'_{ak} = n_{ak} - 1$.

C Dually Consistency of Ordinary MH Estimator

C.1 Sparse Data Limiting Model

For the sparse data limiting model, the number of observations per stratum is bounded ($O(N_k) = 1$) and K approaches infinity.

From $\pi_{j|1k}\pi_{h|2k} = \Psi_{jh}\pi_{h|1k}\pi_{j|2k}$, which follows from the assumption of a common odds ratio, and equation (9), we derive

$$\mathbb{E}\omega_{jh|k} = \mathbb{E}(c_{jh|k} - \Psi_{jh} c_{hj|k}) = \mathbb{E}c_{jh|k} - \Psi_{jh}\mathbb{E}c_{hj|k}$$

$$= \{\mathbb{E}X_{j|1k}\mathbb{E}X_{h|2k} - \Psi_{jh}\mathbb{E}X_{h|1k}\mathbb{E}X_{j|2k}\}/N_k$$

$$= \{n_{1k}n_{2k}\pi_{j|1k}\pi_{h|2k} - \Psi_{jh}n_{1k}n_{2k}\pi_{h|1k}\pi_{j|2k}\}/N_k$$

$$= \{n_{1k}n_{2k}(\pi_{j|1k}\pi_{h|2k} - \pi_{j|1k}\pi_{h|2k})\}/N_k = 0$$

We can write

$$\hat{\Psi}_{jh} - \Psi_{jh} = \frac{\sum_{k=1}^K c_{jh|k} - \Psi_{jh} c_{hj|k}}{\sum_{k=1}^K c_{hj|k}} = \frac{\sum_{k=1}^K (c_{jh|k} - \Psi_{jh} c_{hj|k})/K}{\sum_{k=1}^K c_{hj|k}/K} \tag{12}$$

$$= \frac{\sum_{k=1}^K \omega_{jh|k}/K}{\sum_{k=1}^K c_{hj|k}/K} = \frac{\Omega_{jh}/K}{C_{hj}/K} \tag{13}$$

with with $\omega_{jh|k} := c_{jh|k} - \Psi_{jh} c_{hj|k}$ and $\Omega_{jh} := \sum_k \omega_{jh|k}$.

The term $c_{jh|k}$ is a bounded random variable under model II, hence, the variance of C_{jh} is $o(K^2)$ and Chebyshev's weak law of large numbers states $(\Omega_{jh} - \mathbb{E}\Omega_{jh})/K \to_p 0$. Since $\mathbb{E}\omega_{jh|k} = 0$, the expression $(\Omega_{jh} - \mathbb{E}\Omega_{jh})/K \to_p 0$ reduces to $\Omega_{jh}/K \to_p 0$, that is, the numerator of $\hat{\Psi}_{jh} - \Psi_{jh}$ converges to zero in probability. Applying the Chebyshev weak law of large numbers again to the denominator yields

$$\sum_{k=1}^{K} c_{jh|k}/K \xrightarrow{K \to \infty}_{p} \lim_{K \to \infty} \sum_{k=1}^{K} \mathbb{E}(c_{jh|k})/K < \infty.$$

This limit is finite and nonzero. Thus, we conclude $\hat{\Psi}_{jh} - \Psi_{jh} \to_p 0$ by Slutsky's theorem.

C.2 Large Stratum Limiting Model

Let us consider the case $N \to \infty$ with $N\alpha_{ik} = n_{ik}$ and $0 < \alpha_{ik} < 1$, that is, as N approaches infinity the number of subjects n_{ik}, for all rows i and strata k, also approaches infinity. Note $N_k = n_{1k} + n_{2k} = N \sum_i \alpha_{ik}$.
We have

$$C_{jh}/N = \sum_{k=1}^{K} c_{jh|k}/N = \sum_{k=1}^{K} X_{j|1k} X_{h|2k}/(N_k N)$$

$$= \sum_{k=1}^{K} \frac{n_{1k}n_{2k}}{N_k N} \frac{X_{j|1k}}{n_{1k}} \frac{X_{h|2k}}{n_{2k}} = \sum_{k=1}^{K} \frac{n_{1k}n_{2k}}{NN} \frac{N}{N_k} \frac{X_{j|1k}}{n_{1k}} \frac{X_{h|2k}}{n_{2k}}$$

$$\xrightarrow{N \to \infty}_{p} \sum_{k=1}^{K} \alpha_{1k}\alpha_{2k}(\sum_i \alpha_{ik})^{-1} \pi_{j|1k}\pi_{h|2k} = \sum_{k=1}^{K} (\sum_i \alpha_{ik}^{-1})^{-1} \pi_{j|1k}\pi_{h|2k}.$$

Therefore

$$\hat{\Psi}_{jh} = \frac{C_{jh}}{C_{hj}} = \frac{C_{jh}/N}{C_{hj}/N} \xrightarrow{N \to \infty}_{p} \frac{\sum_{k=1}^{K} (\sum_i \alpha_{ik}^{-1})^{-1} \pi_{j|1k}\pi_{h|2k}}{\sum_{k=1}^{K} (\sum_i \alpha_{ik}^{-1})^{-1} \pi_{h|1k}\pi_{j|2k}}$$

$$= \Psi_{jh} \frac{\sum_{k=1}^{K} (\sum_i \alpha_{ik}^{-1})^{-1} \pi_{h|1k}\pi_{j|2k}}{\sum_{k=1}^{K} (\sum_i \alpha_{ik}^{-1})^{-1} \pi_{h|1k}\pi_{j|2k}} = \Psi_{jh}.$$

D Asymptotic Covariances

D.1 Sparse-data Limiting Model

Let $\text{Var}^a(\cdot)$ and $\text{Cov}^a(\cdot)$ refer to the asymptotic variance and covariance. From above $\hat{\Psi}_{jh} - \Psi_{jh} = \frac{\Omega_{jh}/K}{C_{hj}/K} = \frac{\sum_k \omega_{jh|k}/K}{C_{hj}/K}$.

First by independence of rows $\text{Cov}(\Omega_{jh}, \Omega_{ts}) = \sum_{k=1}^{K} \text{Cov}(\omega_{jh|k}, \omega_{ts|k})$. Note that $\mathbb{E}|\omega_{jh|k} - \mathbb{E}\omega_{jh|k}|^3 = \mathbb{E}|\omega_{jh|k}|^3 = O(1)$, because $c_{jh|k}$ is a bounded random variable under the sparse-data limiting model. By setting $\delta = 1$, we conclude from the Multivariate Central Limit Theorem (Sen and Singer 1993, p.123) that $K^{-1/2}(\Omega_{jh}, \Omega_{ts}) = \sqrt{K}(\Omega_{jh}/K, \Omega_{ts}/K)$ converges to a zero mean multivariate normal distribution with covariance $\lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} \text{Cov}(\omega_{jh|k}, \omega_{ts|k})$, by noting that $\mathbb{E}\omega_{jh|k} = 0$ and $\text{Cov}(\omega_{jh}, \omega_{ts})$ exists. We conclude the asymptotic covariance between Ω_{jh} and Ω_{ts} is $\lim_{K \to \infty} K \cdot \text{Cov}^a(\Omega_{jh}, \Omega_{ts}) = \lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} \text{Cov}(\omega_{jh|k}, \omega_{ts|k})$.

Therefore by the delta method, Slutsky's theorem, equation (12), and using that the denominator terms $\lim_K \mathbb{E}C_{hj}/K$ are finite we obtain

$$\lim_{K \to \infty} K \cdot \mathrm{Cov}^a(\log \hat{\Psi}_{jh}, \log \hat{\Psi}_{ts})$$

$$= 1/(\Psi_{jh}\Psi_{ts}) \lim_{K \to \infty} K \cdot \mathrm{Cov}^a(\hat{\Psi}_{jh}, \hat{\Psi}_{ts})$$

$$= 1/(\Psi_{jh}\Psi_{ts}) \frac{\lim_{K \to \infty} K \cdot \mathrm{Cov}^a(\Omega_{jh}, \Omega_{ts})}{(\lim_K \mathbb{E}C_{hj}/K)(\lim_K \mathbb{E}C_{st}/K)}$$

$$= 1/(\Psi_{jh}\Psi_{ts}) \frac{\lim_{K \to \infty} 1/K \cdot \sum_k \mathrm{Cov}(\omega_{jh|k}, \omega_{ts|k})}{(\lim_K \mathbb{E}C_{hj}/K)(\lim_K \mathbb{E}C_{st}/K)}$$

for arbitrary indices $j, h, s, t \in \{1, \ldots, c\}$ with $j \neq h$ and $s \neq t$.

Now we obtain the following variance

$$\mathrm{Var}(\omega_{jh|k}) = v^1_{jh|k} - 2\Psi_{jh}v^2_{jh|k} + \Psi^2_{jh}v^3_{jh|k}$$

and covariances

$$\mathrm{Cov}(\omega_{jh|k}, \omega_{js|k}) = v_{jhs|12,k} - \Psi_{jh}v_{jh,js|k} - \Psi_{js}v_{js,jh|k} + \Psi_{jh}\Psi_{js}v_{jhs|21,k}$$

$$\mathrm{Cov}(\omega_{jh|k}, \omega_{ts|k}) = v_{jt,hs|k} - \Psi_{jh}v_{ht,js|k} - \Psi_{ts}v_{js,ht|k} + \Psi_{jh}\Psi_{ts}v_{hs,st|k}$$

with

$$v^1_{jh|k} = \frac{n_1 n_2}{N^2}(\pi_{j|1}\pi_{h|2} + n'_1 \pi^2_{j|1}\pi_{h|2} + n'_2 \pi_{j|1}\pi^2_{h|2})$$

$$v^2_{jh|k} = \frac{n_1 n_2}{N^2}(n'_1 \pi_{j|1}\pi_{h|1}\pi_{h|2} + n'_2 \pi_{j|2}\pi_{h|2}\pi_{jh|1} + \pi_{jh|1}\pi_{jh|2})$$

$$v^3_{jh|k} = \frac{n_1 n_2}{N^2}(\pi_{h|1}\pi_{j|2} + n'_1 \pi^2_{h|1}\pi_{j|2} + n'_2 \pi_{h|1}\pi^2_{j|2})$$

$$v_{jh,ts|k} = \frac{n_1 n_2}{N^2}(\pi_{jh|1}\pi_{ts|2} + n'_1 \pi_{j|1}\pi_{h|1}\pi_{ts|2} + n'_2 \pi_{jh|1}\pi_{t|2}\pi_{s|2})$$

$$v_{jhs|abk} = \frac{n_1 n_2}{N^2} v^A_{jhs|abk} + v^B_{jhs|abk} \ (a \neq b)$$

$$v^A_{jhs|abk} = \frac{n_1 n_2}{N^2} \pi_{hs|bk}(\pi_{j|ak} + n'_a \pi^2_{j|ak}) \ (a \neq b)$$

$$v^B_{jhs|abk} = \frac{n_1 n_2}{N^2} n'_b \pi_{j|ak}\pi_{h|bk}\pi_{s|bk} \ (a \neq b).$$

The subscript k is often suppressed for convenience only.

The (co)variance estimators were constructed in such a way that they converge exactly to the asymptotic (co)variance(s). We can also express U_{jhs} as $U_{jhs} = U^{add}_{jhs}$ omitting U^{old}_{jhs} but only if $\hat{v}^B_{jhs|abk}$ is amended to $\hat{v}^B_{jhs|abk} = \frac{1}{N^2_k}X_{j|ak}\{X_{h|bk}X_{s|bk} - X_{hs|bk}\}$. Then for the covariance estimators we have $\sum_k \hat{v}_k/K \overset{K \to \infty}{\to} \sum_k \mathbb{E}\hat{v}_k/K = \lim_K \sum_k v_k/K$ and $\sum_k c_{jh|k}/K \overset{K \to \infty}{\to} \sum_k \mathbb{E}c_{jh|k}/K$ by Chebyshev's weak law of large numbers.

D.2 Large-stratum Limiting Model

By the delta method, the large stratum limiting variance is

$$\lim_{N \to \infty} N \cdot \mathrm{Var}^a(\log \hat{\Psi}_{jh})$$

$$= \frac{\sum_k \frac{\alpha_1^2 \alpha_2}{(\sum_i \alpha_{ik})^2} \{\pi_{j|1}^2 \pi_{h|2} + \Psi_{jh}^2 \pi_{h|1}^2 \pi_{j|2} - 2\Psi_{jh}\pi_{j|1}\pi_{h|1}\pi_{jh|2}\}}{(\sum_k (\sum_i \alpha_{ik}^{-1})^{-1} \pi_{h|1k} \pi_{j|2k})^2}$$

$$+ \frac{\sum_k \frac{\alpha_1 \alpha_2^2}{(\sum_i \alpha_{ik})^2} \{\pi_{j|1}\pi_{h|2}^2 + \Psi_{jh}^2 \pi_{h|1}\pi_{j|2}^2 - 2\Psi_{jh}\pi_{jh|1}\pi_{j|2}\pi_{h|2}\}}{(\sum_k (\sum_i \alpha_{ik}^{-1})^{-1} \pi_{h|1k} \pi_{j|2k})^2}$$

and the limiting covariances are

$$\lim_{N \to \infty} N \cdot \mathrm{Cov}^a(\log \hat{\Psi}_{jh}, \log \hat{\Psi}_{js})$$

$$= \frac{\sum_k \frac{\alpha_1^2 \alpha_2}{(\sum_i \alpha_{ik})^2} \{\pi_{j|1}^2 \pi_{hs|2} - \Psi_{jh}\pi_{j|1}\pi_{h|1}\pi_{js|2} - \Psi_{js}\pi_{j|1}\pi_{s|1}\pi_{jh|2} + \Psi_{jh}\Psi_{js}\pi_{h|1}\pi_{s|1}\pi_{j|2}\}}{(\sum_k (\sum_i \alpha_{ik}^{-1})^{-1} \pi_{h|1k} \pi_{j|2k})^2}$$

$$+ \frac{\sum_k \frac{\alpha_1 \alpha_2^2}{(\sum_i \alpha_{ik})^2} \{\pi_{j|1}\pi_{h|2}\pi_{s|2} - \Psi_{jh}\pi_{jh|1}\pi_{j|2}\pi_{s|2} - \Psi_{js}\pi_{js|1}\pi_{j|2}\pi_{h|2} + \Psi_{jh}\Psi_{js}\pi_{hs|1}\pi_{j|2}^2\}}{(\sum_k (\sum_i \alpha_{ik}^{-1})^{-1} \pi_{h|1k} \pi_{j|2k})^2}$$

$$\lim_{N \to \infty} N \cdot \mathrm{Cov}^a(\log \hat{\Psi}_{jh}, \log \hat{\Psi}_{ts})$$

$$= \frac{\sum_k \frac{\alpha_1^2 \alpha_2}{(\sum_i \alpha_{ik})^2} \{\pi_{j|1}\pi_{t|1}\pi_{hs|2} - \Psi_{jh}\pi_{h|1}\pi_{t|1}\pi_{js|2} - \Psi_{ts}\pi_{j|1}\pi_{s|1}\pi_{ht|2} + \Psi_{jh}\Psi_{ts}\pi_{h|1}\pi_{s|1}\pi_{jt|2}\}}{(\sum_k (\sum_i \alpha_{ik}^{-1})^{-1} \pi_{h|1k} \pi_{j|2k})^2}$$

$$+ \frac{\sum_k \frac{\alpha_1 \alpha_2^2}{(\sum_i \alpha_{ik})^2} \{\pi_{jt|1}\pi_{h|2}\pi_{s|2} - \Psi_{jh}\pi_{ht|1}\pi_{j|2}\pi_{s|2} - \Psi_{ts}\pi_{js|1}\pi_{h|2}\pi_{t|2} + \Psi_{jh}\Psi_{ts}\pi_{hs|1}\pi_{j|2}\pi_{t|2}\}}{(\sum_k (\sum_i \alpha_{ik}^{-1})^{-1} \pi_{h|1k} \pi_{j|2k})^2}.$$

The estimators were constructed such that

$$\lim_{N \to \infty} N \cdot \mathrm{Var}^a(\log \hat{\Psi}_{jh}) = \lim_N N \cdot U_{jhh}$$

$$\lim_{N \to \infty} N \cdot \mathrm{Cov}^a(\log \hat{\Psi}_{jh}, \log \hat{\Psi}_{js}) = \lim_{N \to \infty} N \cdot U_{jhs}$$

$$\lim_{N \to \infty} N \cdot \mathrm{Cov}^a(\log \hat{\Psi}_{jh}, \log \hat{\Psi}_{ts}) = \lim_{N \to \infty} N \cdot U_{jhts}.$$