Estimation and Testing for Association with Multiple-Response Categorical Variables from Complex Surveys

Christopher R. Bilder[1], Thomas M. Loughin[2]
Department of Statistics, University of Nebraska-Lincoln, chris@chrisbilder.com, http://www.chrisbilder.com[1]
Department of Statistics and Actuarial Science, Simon Fraser University Surrey, Surrey, BC, V3T0A3, tloughin@sfu.ca[2]

KEY WORDS: Correlated binary data; Loglinear model; NHANES; Pearson statistic; Pick any/c; Rao-Scott adjustments

1. Introduction

Many surveys include questions that invite respondents to "choose all that apply" or "pick any" from a series of items. A recent Yahoo! internet search on the phrases "survey" and "choose all that apply" yielded over 2000 hits, many from online surveys. Issues of statistical validity of voluntary-response surveys notwithstanding, this points to the fact that these questions are ubiquitous in modern survey methodology. Furthermore, the United States Office of Management and Budget has mandated that federal surveys ask questions about race and ethnicity in a "choose all that apply" format (Federal Register, 1997, p. 58781), allowing members of the increasingly multiracial population to acknowledge their complex ethnicity. Given the frequency with which these questions occur, it is vital to have good methods for statistical analysis of the data they provide.

Variables that summarize survey data arising from "choose all that apply" questions are referred to as multiple-response categorical variables (MRCVs). They present a challenge because they cannot be handled in the same manner as the usual single-response categorical variables (SRCVs), although this has only recently been recognized in the literature (Umesh, 1995; Loughin and Scherer, 1998). When two or more categorical variables are measured, questions naturally arise regarding the associations among them. The difficulty with the analysis of associations involving MRCVs comes from the fact that individual subjects may respond positively to more than one item from a list, and these responses are likely to be correlated. The result is that tests for independence between categorical variables involving MRCVs cannot be performed in the "usual" ways. For example, the Pearson test statistic for independence is not invariant to the arbitrary choice of whether the positive or the negative responses are tabulated (Agresti and Liu, 1999; Bilder, Loughin, and Nettleton, 2000; Bilder and Loughin, 2001). Versions of the Pearson statistic that are modified to be invariant to this choice of coding have distributions that are not, in general, chi-square (Agresti and Liu, 1999; Bilder et al. 2000, Bilder and Loughin, 2004).

Ignoring these problems and simply using the usual Pearson statistic and its associated chi-square distribution provides a test with very poor properties (Loughin and Scherer, 1998). Only recently has work been done to address these problems. For tests of independence between one SRCV and one MRCV, Agresti and Liu (1999), Bilder et al. (2000), Decady and Thomas (2000), and Bilder and Loughin (2001) describe various adjustments to the Pearson statistic and methods to approximate the resulting sampling distributions. Agresti and Liu (1999) point out that MRCVs can be expressed as binary vectors wherein each element of the vector indicates whether the corresponding item is chosen as one of the responses. Decady and Thomas (2000) cleverly note the parallel between an application of an adjusted Pearson statistic to MRCVs and the use of the Pearson statistic in non-multinomial sampling structures as studied by Rao and Scott (1981), although the form of the Pearson statistic used by Decady and Thomas is not invariant to the 0/1 coding of the binary vectors (a different value of the statistic results if the elements indicate non-selection of that item). Thomas and Decady (2004) and Bilder and Loughin (2004) discuss tests of independence between two MRCVs.

Beyond testing for association, there have been a few efforts to model associations involving MRCVs. Agresti and Liu (1999, 2001) propose marginal logit models to describe the association between a single MRCV and a single SRCV. They suggest, but do not explore, extensions to multiple MRCVs. Bilder and Loughin (2003, 2007) examine these suggestions and propose their own modeling procedure. They conclude that a generalized loglinear model fit using a marginal estimation procedure is the preferred model due to computational ease, flexibility, and overall performance. Inference using this model makes use of work by Rao and Scott (1984) and Haber (1985). In particular, model fitting uses a "pseudo" maximum likelihood approach based upon an incorrect assumption of a multinomial distribution for the observed marginal counts, and then adjustments are applied to the sampling distributions of the various estimators and test statistics.

All previous work on MRCVs has been conducted under the assumption of simple random sampling with replacement, so that measurements on sampled subjects can be viewed as a set of independent, identically-distributed random variables from some infinite population. No methods are currently available for the common situation of testing and modeling association

among MRCVs based on data arising from complex survey sampling involving, for example, probability proportional to size, stratification, and/or clustering. The present research combines the results of Rao and Scott (1984) and Bilder and Loughin (2007) to extend existing modeling and analysis techniques in order to provide valid analysis of data from complex survey designs. Beyond the measured responses, the only additional information that the proposed methods require is a set of survey weights that permit unbiased estimation of population totals, and a method of calculating the variance of these estimates. General tests for association, like those developed by Loughin and Scherer (1998), Bilder et al. (2000), Thomas and Decady (2004), and Bilder and Loughin (2004), are obtained as goodness-of-fit tests from the proposed models.

The National Health and Nutrition Examination Survey (NHANES) provides an excellent opportunity to apply the proposed methods. This survey is a large, nationwide study of the health and diet of people in the United States and is conducted periodically using a multistage, complex survey sampling design. The 1999-2000 survey contains numerous questions that can be treated as "choose all that apply" questions. The focus here will be on questions asking the respondent about lifetime tobacco use (>100 cigarettes, >20 pipes, >20 cigars, >20 snuff, and >20 chewing tobacco) and types of respiratory problems experienced (cough on most days, bring up phlegm on most days, experienced wheezing in chest, dry cough at night). The exact survey questions and data are available on the NHANES website at www.cdc.gov/nchs/about/major/nhanes/NHANES99_00.htm. Table 1 displays survey-design adjusted proportions for the number of individuals who responded positively to each item. Note that individual survey respondents may be represented in more than one table cell because they could use more than one type of tobacco and/or have more than one type of respiratory problem so that common Pearson chi-square tests for independence or loglinear models accounting for the survey design can not be used. The purpose of this paper is to show how one can simultaneously model and estimate the association structure between MRCVs in this setting.

This paper is organized as follows. Notation and preliminary details are given in Section 2, followed by a detailed description of the model and associated inference procedures. The association between respiratory symptoms and tobacco use from the NHANES data is analyzed in Section 3. Section 4 presents simulation results that assess the performance of the proposed inference procedures. The general applicability of these methods to other settings are discussed in Section 5.

2. Background

2.1 General notation

Let $\mathcal{U}$ denote a population of N units, and let s be a sample of n units selected from $\mathcal{U}$ according to some probability sampling plan with known first-order inclusion probabilities. Let w_u, $u = 1, \ldots, N$, represent survey weights. These may be simply the inverse of the first-order inclusion probabilities, or they may be more complicated to account for nonresponse, post-stratification, and so forth. Assume that these weights are constructed to lead to unbiased estimates of population totals.

To simplify the exposition, consider the case of two MRCVs, $\boldsymbol{Y} = (Y_1, \ldots, Y_I)$ and $\boldsymbol{Z} = (Z_1, \ldots, Z_J)$, where Y_i is the binary response to item i of $\boldsymbol{Y}$ and Z_j is the binary response to item j of $\boldsymbol{Z}$. Let the corresponding observed values in the population be $\boldsymbol{y}_u = (y_{u1}, \ldots, y_{uI})$ and $\boldsymbol{z}_u = (z_{u1}, \ldots, z_{uJ})$ for $u = 1, \ldots, N$. Extensions to more than two MRCVs are discussed in Section 5. Consider subpopulations of $\mathcal{U}$ corresponding to different $(\boldsymbol{y}, \boldsymbol{z})$ combinations with $\mathcal{U}(\boldsymbol{y}, \boldsymbol{z}) = \{u: (\boldsymbol{y}_u, \boldsymbol{z}_u) = (\boldsymbol{y}, \boldsymbol{z})\}$. The population total count for combination $(\boldsymbol{y}, \boldsymbol{z})$ is $N(\boldsymbol{y}, \boldsymbol{z}) = \sum_{u \in \mathcal{U}} \delta\left(u \in \mathcal{U}(\boldsymbol{y}, \boldsymbol{z})\right)$, where $\delta(\cdot)$ is an indicator function. The sample-weighted estimate of the population total count for combination $(\boldsymbol{y}, \boldsymbol{z})$ is $\tilde{N}(\boldsymbol{y}, \boldsymbol{z}) = \sum_{u \in s} w_u \delta\left(u \in \mathcal{U}(\boldsymbol{y}, \boldsymbol{z})\right)$. These counts can be represented as $2^{I+J} \times 1$ vectors, $\boldsymbol{N}$ and $\tilde{\boldsymbol{N}}$. Without loss of generality, assume that the elements of each vector are arranged in lexicographic order according to the binary numerical value of $(\boldsymbol{y}, \boldsymbol{z})$. Thus, element k corresponds to the combination $(\boldsymbol{y}, \boldsymbol{z})$ that is the binary equivalent of the decimal value $k - 1$.

Next, consider the population marginal count for $(y_i = a, z_j = b)$, where $a, b \in \{0, 1\}$, $M_{ab(ij)} = \sum_{u \in \mathcal{U}} \delta(y_{ui} = a, z_{uj} = b)$, which is estimated by $\tilde{M}_{ab(ij)} = \sum_{u \in s} w_u \delta(y_{ui} = a, z_{uj} = b)$. The corresponding population and estimated proportions are $P_{ab(ij)} = M_{ab(ij)}/N$ and $\tilde{P}_{ab(ij)} = \tilde{M}_{ab(ij)} / \tilde{N}$, respectively, where $\tilde{N} = \sum_{u \in s} w_u$. Table 2 shows the estimated proportions for the respiratory symptoms and tobacco use data from NHANES. Note that $M_{ab(ij)} = \boldsymbol{b}'_{ab(ij)} \boldsymbol{N}$ and $\tilde{M}_{ab(ij)} = \boldsymbol{b}'_{ab(ij)} \tilde{\boldsymbol{N}}$ for a suitably-chosen $1 \times 2^{I+J}$ row vector $\boldsymbol{b}'_{ab(ij)}$. The elements of $\boldsymbol{b}_{ab(ij)}$ are $\delta(y_i = a, z_j = b) = 0$ or 1 with order corresponding to the 2^{I+J} ordered $(\boldsymbol{y}, \boldsymbol{z})$ values. Thus, we have the representations $\boldsymbol{M} = \boldsymbol{B}\boldsymbol{N}$ and $\tilde{\boldsymbol{M}} = \boldsymbol{B}\tilde{\boldsymbol{N}}$, where $\boldsymbol{B}$ is the $4IJ \times 2^{I+J}$ matrix $\boldsymbol{B} = \left(\boldsymbol{b}_{00(11)}, \boldsymbol{b}_{01(11)}, \ldots, \boldsymbol{b}_{11(IJ)}\right)'$.

Analysis of categorical data traditionally focuses on estimating and testing "association" between two or more variables. In particular, associations among binary variables are typically defined in terms of odds ratios. Thus, in the present context, it is natural to represent

association between two MRCVs using pairwise odds ratios between different items of the two variables. This results in IJ odds ratios, $\theta_{ij} = M_{00(ij)}M_{11(ij)}/M_{01(ij)}M_{10(ij)}$, which are easily estimated empirically from Table 2. Questions that immediately arise relate to the presence of structure among these pairwise odds ratios. For example, one may wish to find out whether odds ratios relating tobacco use to a particular respiratory problem are similar across all types of tobacco use. Similarly, it may be of interest to see whether using a certain type of tobacco is associated more strongly with some respiratory problems than with others. Therefore, the modeling procedures used in this paper focus on assessing structure among odds ratios between different items of MRCVs. Other definitions of association between MRCVs exist and are discussed in Section 5.

2.2 The marginal generalized loglinear model

The estimated marginal totals, $\tilde{\boldsymbol{M}}$, form an "item-response table" with structure analogous to Table 2 where 2×2 sub-tables summarize counts for pairwise combinations of items. A loglinear model for $\boldsymbol{M}$ relates the log-odds ratios to parameters that can provide a parsimonious summary of the association structure between the two MRCVs. The base structure of this model is the same as the loglinear model of independence in a two-way contingency table. The model is augmented to provide parameters for all IJ 2×2 sub-tables simultaneously:

$$\log(M_{ab(ij)}) = \gamma_{ij} + \eta^{Y}_{a(ij)} + \eta^{Z}_{b(ij)},$$

for $i = 1, \ldots, I$, $j = 1, \ldots, J$, $a = 0, 1$, and $b = 0, 1$. The γ_{ij} are unknown parameters that control the total counts in each sub-table, and $\eta^{Y}_{a(ij)}$ and $\eta^{Z}_{b(ij)}$ are unknown parameters that control the row and column marginal totals, respectively, in sub-table (i,j), $i = 1, \ldots, I$ and $j = 1, \ldots, J$. As this model contains only parameters for marginal counts in each sub-table, it assumes independence between each $\boldsymbol{Y}$-item and each $\boldsymbol{Z}$-item (all log-odds ratios between items of $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are zero). This structure is referred to by Agresti and Liu (1999, 2001) as "simultaneous pairwise marginal independence" (SPMI). Additional parameters can be added to this model to allow for associations between items of $\boldsymbol{Y}$ and $\boldsymbol{Z}$.

There are IJ odds ratios to be modeled, and these are laid out in a factorial arrangement. Bilder and Loughin (2007) propose adding association parameters in an ANOVA-like fashion – a constant, main effects for each factor, and interactions between factors – that relate the odds ratios to the items of the MRCVs in a structured way. For example, adding a λ_{ab} term that is constant across all i and j results in a "homogeneous association" model with the same odds ratio for each sub-table. Additional terms can be added that allow the log-odds ratios to vary due to $\boldsymbol{Y}$ and/or $\boldsymbol{Z}$ main effects. Section 3 provides examples where these types of model formulations are used. Alternative parameterizations for the association structure can be considered as needed for the application.

Regardless of the parameterization, the generalized loglinear model can be written as $\log(\boldsymbol{M}) = \boldsymbol{X}\boldsymbol{\beta}$ where $\boldsymbol{X} = (\boldsymbol{x}'_{00(11)}, \boldsymbol{x}'_{01(11)}, \ldots, \boldsymbol{x}'_{11(IJ)})'$ is a suitably-chosen $4IJ\times r$ design matrix, $\boldsymbol{\beta}$ is a $r\times1$ vector of parameters, and $\log(\boldsymbol{M})$ is the vector whose elements are $\log(M_{ab(ij)}) = \boldsymbol{x}_{ab(ij)}\boldsymbol{\beta}$. Similar to Haber (1985) and Bilder and Loughin (2007), the model is fit directly to the estimated population marginal totals, $\tilde{\boldsymbol{M}}$, as if they were actually multinomial counts, and adjustments are made to inference procedures to account for the failure of this assumption due to both the sampling design and the correlation of responses to different items by the same respondents. The estimating equations are $\boldsymbol{X}'\tilde{\boldsymbol{M}} = \boldsymbol{X}'\hat{\boldsymbol{M}}$. The model-predicted cell counts in the item-response table are denoted by $\hat{M}_{ab(ij)} = \exp(\boldsymbol{x}_{ab(ij)}\hat{\boldsymbol{\beta}})$ where $\hat{\boldsymbol{\beta}}$ is the estimated parameter vector. The model-fitting procedure is generally robust against sparse data among the $(\boldsymbol{y}, \boldsymbol{z})$ response combination vectors. This is important, as sparseness is a problem that plagues other approaches to modeling MRCVs (see Bilder et al., 2000 and Bilder and Loughin, 2007).

2.3 Model-comparison statistics

Let $\hat{M}^{(0)}_{ab(ij)}$ and $\hat{M}^{(1)}_{ab(ij)}$ be the model-predicted population totals under some null and alternative hypothesis models, respectively, and assume the null hypothesis model is nested within the alternative model. Similarly, let $\hat{P}^{(0)}_{ab(ij)}$ and $\hat{P}^{(1)}_{ab(ij)}$ be the model-based estimated population probabilities. A Pearson statistic to compare the two models is

$$X^2 = n\sum_{i,j,a,b}\frac{\left(\hat{P}^{(1)}_{ab(ij)} - \hat{P}^{(0)}_{ab(ij)}\right)^2}{\hat{P}^{(0)}_{ab(ij)}} = \frac{n}{\tilde{N}}\sum_{i,j,a,b}\frac{\left(\hat{M}^{(1)}_{ab(ij)} - \hat{M}^{(0)}_{ab(ij)}\right)^2}{\hat{M}^{(0)}_{ab(ij)}}.$$

Because the item-response table counts do not have a multinomial distribution and data arises under complex survey sampling, X^2 does not have an asymptotic chi-square distribution. Instead, Appendix A shows that the asymptotic distribution is a linear combination of independent χ^2_1 random variables.

For analyzing associations between SRCVs from complex sampling designs, Rao and Scott (1984) propose first- and second-order adjustments to a statistic like X^2 that allow approximate inference to be based on a single chi-square distribution. The first-order adjustment matches the mean of the test statistic to the mean of the

reference distribution. Applying this adjustment to X^2 here results in a test statistic of the form $X^2_{RS1} = HX^2 / \sum_{h=1}^{H} \hat{\gamma}_h$, where $\hat{\gamma}_h$, for $h = 1, \ldots, H$, are the coefficients in the linear combination for the asymptotic distribution of X^2. This statistic is judged against a χ^2_H distribution. The second-order adjustment matches both the means and the variances of the test statistic and the reference distribution. This leads here to a test statistic $X^2_{RS2} = (\sum_{h=1}^{H} \hat{\gamma}_h) X^2 / \sum_{h=1}^{H} \hat{\gamma}_h^2$, which is compared to a χ^2_ν distribution with $\nu = (\sum_{h=1}^{H} \hat{\gamma}_h)^2 / \sum_{h=1}^{H} \hat{\gamma}_h^2$.

Also for SRCVs from complex sampling designs, Thomas, Singh, and Roberts (1996) show that an F-distribution approximation to a further modified version of X^2 generally holds the correct size of the test better. Following Rao and Thomas (2003) and incorporating this into the MRCV setting, the adjusted statistic is $F = X^2_{RS1} / (H^{-1} \sum_{h=1}^{H} \hat{\gamma}_h)$. The F statistic can be approximated by an $F_{H,H\nu}$ random variable for a first-order adjustment where ν is the degrees of freedom resulting from the estimation of $Cov(\tilde{\boldsymbol{M}})$ (for example, $\nu = 4IJ - 1$ under probability-proportional-to-size sampling). For a second-order adjustment, F can be approximated by an $F_{H^*,H^*\nu}$ random variable where $H^* = H/(1+\hat{a}^2)$ and $\hat{a}^2 = \left[H \sum_{h=1}^{H} \hat{\gamma}_h^2 / (\sum_{h=1}^{H} \hat{\gamma}_h)^2 \right] - 1$.

Measures of deviations from a model can be found through standardized Pearson residuals. Using the corresponding diagonal element of the asymptotic covariance matrix for the residuals in Appendix A, the standardized Pearson residual is

$$(\tilde{M}_{ab(ij)} - \hat{M}_{ab(ij)}) \Big/ \sqrt{\hat{AsVar}(\tilde{M}_{ab(ij)} - \hat{M}_{ab(ij)})}.$$

Once a model has been found that fits adequately, model-based odds ratios can be used to interpret the association between items of different MRCVs. Specifically, a model-based estimated odds ratio for the (i, j) sub-table is $\hat{\theta}_{ij} = \hat{M}_{11(ij)} \hat{M}_{00(ij)} / \hat{M}_{01(ij)} \hat{M}_{10(ij)}$. Confidence intervals for the true odds ratio can also be found from the model using asymptotic normality for the estimator and the corresponding standard error for $\log(\hat{\theta}_{ij})$ as derived in Appendix A.

3. Application to the NHANES data

The target population for the lifetime tobacco use section of the NHANES was United States residents age 20 and older. In addition to survey weights that are available for each individual in the NHANES, fifty-two jackknife replicates are provided to aid in the estimation of variances and covariances. For instance, the estimated covariance matrix for $\tilde{\boldsymbol{N}}$ is

$$\tilde{\boldsymbol{V}} = \frac{(52-1)^2}{52^2} \sum_{\ell=1}^{52} \left(\tilde{\boldsymbol{N}}_{(\ell)} - \tilde{\boldsymbol{N}} \right) \left(\tilde{\boldsymbol{N}}_{(\ell)} - \tilde{\boldsymbol{N}} \right)'$$

where $\tilde{\boldsymbol{N}}_{(\ell)}$ is a vector of population total estimates that excludes the ℓ^{th} replicate.

A variety of models are fit providing different descriptions of the association structure among items between the two MRCVs. Table 3 summarizes the goodness-of-fit statistics comparing these models to a saturated model. Clearly, independence does not hold in these sub-tables. Simply adding λ_{ab} to the SPMI model allows for homogenous association across sub-tables and results in a model that fits the data reasonably well. Adding main effects for tobacco use and/or respiratory symptoms result in only slight improvements relative to the homogeneous association model, although the model that includes both main effects clearly fits well relative to a saturated model.

Standardized Pearson residuals for the homogenous association and tobacco and respiratory main effects models are given in Table 4. Note that the absolute value of the standardized Pearson residuals is the same within each cell of a 2×2 sub-table. As in the analysis of ordinary loglinear models, these residuals have approximate standard normal distributions when the fitted model is correct. The homogenous association model does have a few values a little larger than expected, but the tobacco and respiratory main effects model does not. For this reason, the tobacco and respiratory main effects model is chosen for further investigation of the data.

Restricting attention to ANOVA-type models is convenient, but by no means necessary. For example, notice that there is a stronger association between cigarette use and the first two respiratory problems than between any other combination of tobacco use and respiratory problems. When one parameter is added to the homogeneous association model indicating whether or not a count is from one of these two sub-tables, X^2_{RS2} results in a goodness-of-fit p-value of 0.6150. The largest standardized Pearson residual is 1.98 in absolute value. Because this model is not nested within the tobacco and respiratory main effects model, a hypothesis test is not easily performed to compare them. Of course the usual caveats apply for testing hypotheses that are suggested by the data, but the point is that flexibility exists within these models to describe a wide variety of potential association structures.

Returning to the tobacco and respiratory main effects model, Table 4 provides the model-estimated odds ratios and corresponding 95% confidence intervals. Also, the table gives the empirical odds ratios, $\tilde{\theta}_{emp,ij} = \tilde{M}_{11(ij)} \tilde{M}_{00(ij)} / \tilde{M}_{01(ij)} \tilde{M}_{10(ij)}$, and the corresponding 95% confidence intervals using the asymptotic normality of the estimator (see Appendix A for derivation). All odds-ratio confidence intervals are above one indicating a positive association between each type of tobacco use and the presence of each respiratory problem. Overall, the strongest estimated association appears to be with cigarette use and the respiratory problems. This

information may be useful then for public health purposes and treatment programs.

4. Simulation study

The actual NHANES sampling design is quite complicated, so it would be extremely difficult to duplicate exactly in a simulation study. Instead, a simpler sampling design – specifically, unequal probability sampling – is used here for a simulation study in order to evaluate the performance of the proposed model comparison statistics and odds ratio inference procedures in a controlled setting.

Motivated by Rai, Srivastava, and Gupta (2001), the general approach to the simulations is as follows. Simulation settings are first chosen and each is used to create a population with certain known characteristics. Specifically, a vector of multinomial joint probabilities is constructed corresponding to all 2^{I+J} possible binary response vectors, (***y***, ***z***). These probabilities are chosen to reflect various marginal dependence structures between items of different MRCVs and within items of each MRCV. Past research involving MRCVs under simple random sampling have indicated that the level of dependence within the MRCVs can affect the performance of model-comparison tests (see, e.g., Bilder et al., 2000). Two different marginal dependence structures are examined in the simulation study, representing both high and low dependence between items within each MRCV. These marginal specifications are translated into the multinomial joint probabilities using the algorithm of Gange (1995). The population is then formed by simply multiplying the multinomial probabilities by N = 100,000 and rounding to the nearest integer to form counts for the 2^{I+J} possible responses. Due to the rounding, the actual population size may be slightly smaller or larger than 100,000.

Next, unequal probability sampling without replacement is applied to each population for a fixed sample size and using the Hanurav-Vijayan algorithm (Fox (1989), Golmant (1990), and Watts (1991)). Sample sizes range from 50 to 500, representing small to moderately sized surveys. Three different sets of inclusion probabilities were created for each population, representing 1) equal probabilities (simple random sampling), 2) moderate differences in probabilities, and 3) large differences in probabilities. The moderate differences are created by randomly assigning population units into one of five equally sized groups. From there, each unit within a group is assigned the same inclusion probability. These inclusion probabilities are chosen to be proportional to 1, 1.5, 2, 2.5, and 3 leading to the largest inclusion probability being three times the smallest. The large differences are created by assigning the inclusion probability for each unit to be proportional to a Uniform(0,1) simulated value.

For each set of marginal dependence structures and (I, J), one population is created and 500 samples of the same size are taken from it. Simulated data sets with a $\tilde{M}_{ab(ij)} = 0$ are excluded in order to evaluate the procedures only when all model parameters are estimated for each data set. This occurred only a few times for the smaller sample size simulations and excluded at most 19 out of 500 data sets. For each simulated data set used, specified models are fit and model comparison statistics are calculated. A jackknife estimate of $\boldsymbol{V}$,

$$\tilde{\boldsymbol{V}} = \frac{n^2}{(n-1)^2}\sum_{\ell=1}^{n}\left(\tilde{\boldsymbol{N}}_{(\ell)} - \tilde{\boldsymbol{N}}\right)\left(\tilde{\boldsymbol{N}}_{(\ell)} - \tilde{\boldsymbol{N}}\right)'$$

where $\tilde{\boldsymbol{N}}_{(\ell)}$ are estimated population totals for a jackknife resample that excludes the ℓ^{th} observation, is found for each data set to estimate the covariance matrix of $\tilde{\boldsymbol{N}}$. This matrix is used in calculations needed for the distributional approximations with model comparison statistics and odds ratios (see Appendix A). A significance level of 0.05 is used throughout for inference procedures. The approximate 95% expected range for the estimated size of tests is $0.05 \pm 1.96[0.05(1\text{-}0.05)//500]^{1/2}$ = (0.031, 0.069).

Table 5 summarizes the estimated size of tests when the population satisfies the important special case of SPMI and two-way interactions between items within the same MRCV are controlled at the levels specified in the table. The model under SPMI is fit to the simulated data sets and the corresponding goodness-of-fit statistic is found. As seen from the table, the first-order adjusted statistics reject too often when strong pairwise association exists between items within an MRCV. The second-order adjusted statistics generally hold the correct size, but can be a little conservative at times, especially for F. All of these results hold over the different sampling plans.

Additional simulations were also performed and led to similar results. For example, model-comparison tests were examined for populations created under homogenous association (all sub-table odds ratios are equal) and ***Y***-main effects (heterogeneity of sub-table odds ratios across levels of ***Y***). The patterns of rejection rates for the various tests are similar to those observed here under SPMI. Also, data were simulated so that a three-way interaction could be imposed among the items within ***Z***. Again, very similar results are found in this setting.

When SPMI is not satisfied in the population, it is of interest to examine how well the models estimate odds ratios between items of different MRCVs. For two types of deviations from SPMI, Table 6 provides the bias and mean square error for the point estimates and the coverage and mean length for 95% confidence intervals.

These summary measures are averaged over sub-tables with equal θ_{ij} values to facilitate direct comparisons between empirical-based and model-based measures. Overall, the model-based procedure's mean square errors for estimating odds ratios are uniformly smaller than those from the empirical odds ratio and its confidence interval mean lengths are up to 50% shorter under the observed conditions. The biases are similar and coverage levels are close to the nominal level for both procedures. These results hold over the different sampling plans, but the mean square error and confidence interval mean length increase as the variability among the inclusion probabilities increases. Additional simulations performed under other settings showed similar results.

5. Discussion

The simulations show that there is a distinct advantage to modeling associations in MRCVs rather than simply analyzing associations through a series of isolated 2×2 contingency tables. Tests for interesting structures are available with the models, and the variability of the estimates is reduced due to the simultaneous estimation.

It is important to note that the applicability of these methods reaches beyond the scope of questions asked directly as "choose all that apply" or "pick any" from a set of item responses. Because an MRCV is merely a special kind of correlated binary data vector, the statistical analysis methods that have been developed in this paper can be applied much more broadly to certain marginal summaries of any sequence of binary random variables. In particular, series of "yes/no" questions from surveys that cover closely related topics, but are not necessarily posed as "choose all that apply", can be combined into a binary vector to form a MRCV and modeled along with other such vectors. In fact, both questions of the example analyzed in Section 3 were presented in this format. Additional examples of surveys containing questions of this form include other NHANES questionnaires, the Harvard School of Public Health College Alcohol Study, the U.S. Census Bureau and Bureau of Labor Statistics' Current Population Survey, and the National Center for Health Statistics' National Survey of Family Growth.

The models discussed here also may be used when there are more than two MRCVs present. Bilder and Loughin (2007) show how these models are used in the simple random sampling case. The main idea is to construct sub-tables for each combination of items from different MRCVs. For example, to model three MRCVs, there will be IJK different $2\times2\times2$ sub-tables summarizing counts within the item-response table, where K is the number of items for a third MRCV, say $\boldsymbol{W}$. The base model representing complete marginal independence starts with parameters for the overall sub-table count and each one-dimensional margin of every $2\times2\times2$ sub-table. Additional parameters are added as needed to specify structures for the three forms of two-way association ($\boldsymbol{WY}$, $\boldsymbol{WZ}$, and $\boldsymbol{WZ}$) and the three-way association. Asymptotic distributions for model comparison statistics, standardized Pearson residuals, and other statistics can be derived in a similar manner as given in Appendix A of this paper. Of course, as the number of MRCVs increase, the item-response table and corresponding models become more complicated. When there is only one MRCV and one SRCV, models similar to those in Agresti and Liu (1999) can be constructed as well. The item-response table in this case would consist of J different $I\times2$ sub-tables.

The models in this paper focus on estimating and interpreting certain *marginal* associations among variables. Other forms of association could be considered instead. For example, Berry and Mielke (2003) discuss a permutation approach to testing for *joint* independence between multinomial vectors $\boldsymbol{y}$ and $\boldsymbol{z}$ in the case of simple random sampling with replacement. Testing for joint independence reduces to testing for independence in the $2^I\times2^J$ table of $\boldsymbol{y}\times\boldsymbol{z}$ responses, which results in testing for a much stronger form of independence than SPMI. The $2^I\times2^J$ table is likely to be very sparse if I and J are not both very small, so that parametric modeling methods are likely to be difficult to adopt. Also, extracting interesting and interpretable patterns of joint association is considerably more difficulty because summary measures of association in large two-way contingency tables are not nearly as well understood or as easily interpreted as the odds ratios used here to represent the marginal associations.

When estimating parameters from ordinary loglinear models for SRCV data under complex survey sampling, Clogg and Eliason (1987) and Magidson (1987) recommend fitting a conditional rate model to the observed counts (not adjusted for the survey design) while using an average survey weight for each table cell as an offset. Hendrickx (2002) examines this model formulation along with those from Rao and Scott (1984) and shows for a number of examples that the conditional rate model produces smaller standard errors for parameter estimates. While smaller standard errors are desirable, they are a few problems with averaging survey weights over individual sampling units and treating them as constants. First, Patterson, Dayton, and Graubard (2002) choose not to use a conditional rate approach in a latent class analysis model due to its inability to account for clustering in a survey design. Second, the standard errors can be underestimated if the variability in the survey weights is not taken into account. For the MRCV problem here, a conditional rate model formulation of the

model was investigated in the simulations. Consistently, the model comparison statistics (with first and second-order adjustments to account for the MRCVs) had inflated type I error rates for sampling with moderate and large differences in probabilities. Confidence intervals for odds ratios had inadequate coverage. These problems were more pronounced as the variability in the inclusion probabilities increased. Due to these problems, we do not recommend the conditional rate model for modeling MRCVs.

Acknowledgements

This work was supported in part by National Science Foundation grants SES-0418688 and SES-0418632. The authors thank Randy Sitter for numerous helpful suggestions.

References

Agresti, A. (2002). *Categorical Data Analysis*. New York: John Wiley & Sons.

Agresti, A. and Liu, I.-M. (1999). Modeling a categorical variable allowing arbitrarily many category choices. *Biometrics* 55, 936-943.

Agresti, A. and Liu, I.-M. (2001). Strategies for modeling a categorical variable allowing multiple category choices. *Sociological Methods & Research* 29, 403-434.

Berry, K. J. and Mielke, P. W. (2003). Permutation analysis of data with multiple binary category choices. *Psychological Reports* 92, 91-98.

Bilder, C. R., Loughin, T. M., and Nettleton, D. (2000). Multiple marginal independence testing for pick any/c variables. *Communications in Statistics: Simulation and Computation* 29, 1285-1316.

Bilder, C. R. and Loughin, T. M. (2001). On the first-order Rao-Scott correction of the Umesh-Loughin-Scherer statistic. *Biometrics* 57, 1253-1255.

Bilder, C. R. and Loughin, T. M. (2002). Testing for conditional multiple marginal independence. *Biometrics* 58, 200-208.

Bilder, C. R. and Loughin, T. M. (2003). Strategies for modeling two categorical variables with multiple category choices. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 560-567.

Bilder, C. R. and Loughin, T. M. (2004). Testing for marginal independence between two categorical variables with multiple responses. *Biometrics* 60, 241-248.

Bilder, C. R. and Loughin, T. M. (2007). Modeling association between two or more categorical variables that allow for multiple category choices. *Communications in Statistics: Theory and Methods* 36, 433-451.

Clogg, C. C., and Eliason, S. R. (1987). Some common problems in log-linear analysis. *Sociological Methods & Research* 16, 8–44.

Decady, Y. J. and Thomas, D. H. (2000). A simple test of association for contingency tables with multiple column responses. *Biometrics* 56, 893-896.

Federal Register (1997). 62, 58781-58790 (October 30).

Fox, D. R. (1989). Computer selection of size-biased samples. *The American Statistician* 43, 168-171.

Gange, S. J. (1995). Generating multivariate categorical variates using the iterative proportional fitting algorithm. *The American Statistician* 49, 134-138.

Golmant, J. (1990). Correction: Computer selection of size-biased samples. *The American Statistician* 44, 194.

Haber, M. (1985). Log-linear models for correlated marginal totals of a contingency table. *Communications in Statistics: Theory and Methods* 14, 2845-2856.

Hendrickx, J. (2002). The impact of weights on standard errors. "Targeting Mr. X - but is he Mr. Right: Sampling, Weighting, Profiling, Segmentation, and Modelling," conference on 4/17/02 at Imperial College (available online at http://www.asc.org.uk/Events/ Apr02/ Full/Hendrickx.doc).

Loughin, T. M. and Scherer, P. N. (1998). Testing for association in contingency tables with multiple column responses. *Biometrics* 54, 630-637.

Magidson, J. (1987). Weighted log-linear modeling. *American Statistical Association Proceedings of the Section on Social Statistics*, 171-175.

Patterson, B. H., Dayton, C. M., and Graubard, B. I. (2002). Latent class analysis of complex sample survey data: application to dietary data. *Journal of the American Statistical Association* 97, 721-741.

Rai, A., Srivastava, A. K., and Gupta, H. C. (2001). Small sample comparison of modified chi-square test statistics for survey data. *Biometrical Journal* 43, 483-495.

Rao, J. N. K. and Scott, A. J. (1981). The analysis of categorical data from complex surveys: chi-squared tests for goodness of fit and independence in two-way tables. *Journal of the American Statistical Association* 76, 221-230.

Rao, J. N. K. and Scott, A. J. (1984). On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *Annals of Statistics* 12, 46-60.

Rao, J. N. K. and Thomas, D. R. (2003). Analysis of categorical response data from complex surveys: an appraisal and update. In Chambers, R. L. and Skinner, C. J. (Eds.), *Analysis of Survey Data* (p. 85 – 108). John Wiley & Sons: New York.

Thomas, D. R. and Decady, Y. J. (2004). Testing for association using multiple response survey data: approximate procedures based on the Rao-Scott approach. *International Journal of Testing* 4, 43-59.

Thomas, D. R., Singh, A. C., and Roberts, G. R. (1996). Tests of independence on two-way tables under cluster sampling: an evaluation. *International Statistical Review* 64, 295-311.

Umesh, U. N. (1995). Predicting nominal variable relationships with multiple responses. *Journal of Forecasting* 14, 585-596.

Watts, D. L. (1991). Correction: computer selection of size-biased samples. *The American Statistician* 45, 172.

Appendix A

The general approach to deriving asymptotic distributions for model parameter estimates and associated quantities is similar to those used in Agresti (2002, Chapter 14) and Rao and Scott (1984). Let $\boldsymbol{V} = Cov(\tilde{\boldsymbol{N}})$, and assume that some consistent estimate, $\tilde{\boldsymbol{V}}$, can be obtained. For example, Sections 3 and 4 show how a jackknife estimator can be found. The asymptotic distribution of the model parameter estimator can be derived from the fact that $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \approx (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{D}_{\boldsymbol{M}}^{-1/2}\boldsymbol{B}(\tilde{\boldsymbol{N}} - \boldsymbol{N})$ where $\boldsymbol{\beta}_0$ is the true value of $\boldsymbol{\beta}$, $\boldsymbol{D}_{\boldsymbol{M}} = \text{Diag}(\boldsymbol{M})$, and

$$\boldsymbol{A} = \boldsymbol{D}_{\boldsymbol{M}}^{-1/2}\left.\frac{\partial \boldsymbol{M}}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$$

where $\boldsymbol{M} = \exp(\boldsymbol{X}\boldsymbol{\beta})$. Noting that

$$\frac{\partial \boldsymbol{M}}{\partial \boldsymbol{\beta}} = \boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{X},$$

we see then

$$\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \dot{\sim} N\left(\boldsymbol{0}, \left(\boldsymbol{X}'\boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{B}\boldsymbol{V}\boldsymbol{B}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{X}\right)^{-1}\right).$$

Properties of the model predicted totals are found from

$$\left(\hat{\boldsymbol{M}} - \boldsymbol{M}\right) \approx \left(\left.\frac{\partial \boldsymbol{M}}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$$

resulting in $\left(\hat{\boldsymbol{M}} - \boldsymbol{M}\right) \dot{\sim}$

$$N\left(\boldsymbol{0}, \boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{B}\boldsymbol{V}\boldsymbol{B}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{D}_{\boldsymbol{M}}\right).$$

Inferences about odds ratios are based on the fact that $\log(\theta_{ij}) = \boldsymbol{l}_{ij}'\boldsymbol{\beta}$ where $\boldsymbol{l}_{ij}' = \boldsymbol{x}_{00(ij)} + \boldsymbol{x}_{11(ij)} - \boldsymbol{x}_{10(ij)} - \boldsymbol{x}_{01(ij)}$. Thus, the model-predicted log odd ratio,

$$\log\left(\hat{\theta}_{ij}\right) = \boldsymbol{l}_{ij}'\hat{\boldsymbol{\beta}},$$

has an asymptotic distribution represented by

$$\left(\log\left(\hat{\theta}_{ij}\right) - \log\left(\theta_{ij}\right)\right) \dot{\sim}$$

$$N\left(\boldsymbol{0}, \boldsymbol{l}_{ij}'\left(\boldsymbol{X}'\boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{B}\boldsymbol{V}\boldsymbol{B}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{X}\right)^{-1}\boldsymbol{l}_{ij}\right).$$

The asymptotic distribution of the empirical log-odds ratio can be represented by

$$\left(\log\left(\tilde{\theta}_{emp,ij}\right) - \log\left(\theta_{ij}\right)\right) \dot{\sim} N\left(\boldsymbol{0}, \boldsymbol{g}_{ij}\boldsymbol{B}\boldsymbol{V}\boldsymbol{B}'\boldsymbol{g}_{ij}'\right)$$

where $\tilde{\theta}_{emp,ij} = \log\left(M_{11(ij)}M_{00(ij)} / M_{01(ij)}M_{10(ij)}\right)$ and $\boldsymbol{g}_{ij} = (d / d\boldsymbol{M})\theta_{emp,ij}$.

Standardized Pearson residuals for the cells of the item-response table are defined as

$$r_{ab(ij)} = e_{ab(ij)} \Big/ \sqrt{\hat{As}Var(e_{ab(ij)})}$$

where $e_{ab(ij)} = \tilde{M}_{ab(ij)} - \hat{M}_{ab(ij)}$. A vector of these residuals, $\boldsymbol{e}$, can be written as $(\tilde{\boldsymbol{M}} - \hat{\boldsymbol{M}}) = (\tilde{\boldsymbol{M}} - \boldsymbol{M}) - (\hat{\boldsymbol{M}} - \boldsymbol{M})$. The joint asymptotic distribution of these two quantities is

$$\begin{bmatrix}\tilde{\boldsymbol{M}} - \boldsymbol{M} \\ \hat{\boldsymbol{M}} - \boldsymbol{M}\end{bmatrix} \approx \begin{bmatrix}\boldsymbol{I} \\ \boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\end{bmatrix}\boldsymbol{B}(\tilde{\boldsymbol{N}} - \boldsymbol{N})$$

from which

$$\boldsymbol{e} = \begin{bmatrix}\boldsymbol{I} & -\boldsymbol{I}\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{M}} - \boldsymbol{M} \\ \hat{\boldsymbol{M}} - \boldsymbol{M}\end{bmatrix}$$

$$\approx \left(\boldsymbol{I} - \boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right)\boldsymbol{B}(\tilde{\boldsymbol{N}} - \boldsymbol{N}).$$

Thus, $\boldsymbol{e} \dot{\sim} N(\boldsymbol{0}, \boldsymbol{G}\boldsymbol{B}\boldsymbol{V}\boldsymbol{B}'\boldsymbol{G}')$ where $\boldsymbol{G} = \boldsymbol{I} - \boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{X})^{-1}\boldsymbol{X}'$.

The Pearson statistic can be used to compare models, $\mathcal{M}_0$: $\log(\boldsymbol{M}(\boldsymbol{\beta}_R)) = \boldsymbol{X}_R\boldsymbol{\beta}_R$, the null or reduced model, against model $\mathcal{M}_1$: $\log(\boldsymbol{M}(\boldsymbol{\beta})) = \boldsymbol{X}\boldsymbol{\beta}$, the alternative or full model, where $\boldsymbol{X} = \begin{bmatrix}\boldsymbol{X}_R & \boldsymbol{X}_{F-R}\end{bmatrix}$ and $\boldsymbol{\beta}' = \begin{bmatrix}\boldsymbol{\beta}_R' & \boldsymbol{\beta}_{F-R}'\end{bmatrix}$, with similar notation for the estimated and true values of the parameters. Let the predicted cell totals from $\mathcal{M}_0$ be denoted by $\hat{\boldsymbol{M}}^{(0)}$ and those from $\mathcal{M}_1$ by $\hat{\boldsymbol{M}}^{(1)}$. The Pearson statistic is

$$X^2 = n\tilde{N}^{-1}\left(\hat{\boldsymbol{M}}^{(1)} - \hat{\boldsymbol{M}}^{(0)}\right)' \boldsymbol{D}_{\hat{\boldsymbol{M}}^{(0)}}^{-1}\left(\hat{\boldsymbol{M}}^{(1)} - \hat{\boldsymbol{M}}^{(0)}\right)$$

Through a variety of manipulations involving the inverse of a partitioned matrix, we come to the result that

$$\left(\hat{\boldsymbol{M}}^{(1)} - \hat{\boldsymbol{M}}^{(0)}\right) \approx \boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{Q}\hat{\boldsymbol{\beta}}_{F-R}$$

where $\boldsymbol{Q} = \left[\boldsymbol{I} - \boldsymbol{X}_R(\boldsymbol{X}_R'\boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{X}_R)^{-1}\boldsymbol{X}_R'\boldsymbol{D}_{\boldsymbol{M}}\right]\boldsymbol{X}_{F-R}$. Thus, X^2 can be written approximately as

$$X^2 \approx n\tilde{N}^{-1}\hat{\boldsymbol{\beta}}_{F-R}'\boldsymbol{Q}'\boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{D}_{\hat{\boldsymbol{M}}^{(0)}}^{-1}\boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{Q}\hat{\boldsymbol{\beta}}_{F-R}$$

$$\approx n\tilde{N}^{-1}\hat{\boldsymbol{\beta}}_{F-R}'\boldsymbol{Q}'\boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{Q}\hat{\boldsymbol{\beta}}_{F-R}$$

Under H_o: $\mathcal{M}_0$ holds, we can determine that the asymptotic distribution of X^2 is the same as that of $\sum_{h=1}^{H}\gamma_h\xi_h^2$ where ξ_h^2 are independent χ_1^2 random variables, $H = \text{rank}(\boldsymbol{X}_{F-R})$, and γ_h are the eigenvalues of

$$n\tilde{N}^{-1}\boldsymbol{Q}'\boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{Q}\left[\left(\boldsymbol{Q}'\boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{Q}\right)^{-1}\boldsymbol{Q}'\boldsymbol{B}\boldsymbol{V}\boldsymbol{B}'\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{Q}\right)^{-1}\right]$$

$$= n\tilde{N}^{-1}\left(\boldsymbol{Q}'\boldsymbol{B}\boldsymbol{V}\boldsymbol{B}'\boldsymbol{Q}\right)\left(\boldsymbol{Q}'\boldsymbol{D}_{\boldsymbol{M}}\boldsymbol{Q}\right)^{-1}.$$

In all cases, variances are estimated by replacing $\boldsymbol{M}$ and $\boldsymbol{V}$ with their respective estimates. The estimated eigenvalues are denoted by $\hat{\gamma}_h$ for $h = 1, \ldots, H$.

Table 1. Survey-design adjusted proportions for the NHANES data.

		Respiratory problems during past year			
		Coughing most days during a 3 month period	Bring up phlegm most days during a 3 month period	Wheezing or whistling in chest	Dry cough at night
Lifetime tobacco use	Cigarettes >= 100 times	0.0517	0.0523	0.0962	0.0248
	Pipe >= 20 times	0.0127	0.0134	0.0204	0.0059
	Cigars >= 20 times	0.0168	0.0174	0.0288	0.0092
	Snuff >= 20 times	0.0055	0.0061	0.0129	0.0039
	Chewing tobacco >= 20 times	0.0064	0.0095	0.0137	0.0036

Table 2. Item-response table providing survey-design adjusted proportions for the NHANES data.

			Respiratory problems during past year							
			Coughing most days during a 3 month period		Bring up phlegm most days during a 3 month period		Wheezing or whistling in chest		Dry cough at night	
			0	1	0	1	0	1	0	1
Lifetime tobacco use	Cigarettes >= 100 times	0	0.4860	0.0196	0.4882	0.0175	0.4556	0.0500	0.4931	0.0125
		1	0.4427	0.0517	0.4420	0.0523	0.3982	0.0962	0.4695	0.0248
	Pipe >= 20 times	0	0.8491	0.0586	0.8513	0.0565	0.7819	0.1258	0.8763	0.0315
		1	0.0796	0.0127	0.0789	0.0134	0.0719	0.0204	0.0864	0.0059
	Cigars >= 20 times	0	0.8009	0.0545	0.8029	0.0524	0.7379	0.1174	0.8272	0.0282
		1	0.1278	0.0168	0.1273	0.0174	0.1159	0.0288	0.1355	0.0092
	Snuff >= 20 times	0	0.8806	0.0658	0.8827	0.0637	0.8131	0.1333	0.9130	0.0334
		1	0.0481	0.0055	0.0475	0.0061	0.0407	0.0129	0.0497	0.0039
	Chewing tobacco >= 20 times	0	0.8775	0.0650	0.8822	0.0603	0.8100	0.1325	0.9087	0.0338
		1	0.0511	0.0064	0.0480	0.0095	0.0438	0.0137	0.0539	0.0036

Table 3. Model comparison p-values for various models.

		P-values			
H_o model	H_a model	X^2_{RS1}	X^2_{RS2}	F 1st	F 2nd
SPMI	Saturated	<0.0001	<0.0001	<0.0001	<0.0001
Homogenous association	Saturated	0.0782	0.1575	0.0801	0.1600
Tobacco main effect	Saturated	0.1407	0.2016	0.1429	0.2041
Respiratory main effect	Saturated	0.1614	0.2322	0.1636	0.2346
Tobacco & Respiratory main effects	Saturated	0.5278	0.4886	0.5283	0.4895
SPMI	Homogenous association	<0.0001	<0.0001	<0.0001	<0.0001
Homogenous association	Tobacco main effect	0.1666	0.1877	0.1695	0.1908
Homogenous association	Respiratory main effect	0.1209	0.1327	0.1240	0.1360
Homogenous association	Tobacco & Respiratory main effects	0.0869	0.1195	0.0893	0.1223
Tobacco main effect	Tobacco & Respiratory main effects	0.1104	0.1228	0.1134	0.1260
Respiratory main effect	Tobacco & Respiratory main effects	0.1378	0.1619	0.1407	0.1650

Table 4. Fit and diagnostic measures for the homogenous association (HA) and tobacco and respiratory main effects (TR) models. For each sub-table, $\hat{\theta}_{HA,ij}$ = 2.11 and the 95% confidence interval for θ_{ij} is (1.77, 2.50) for the HA model. Highlighted cells correspond to $|r_{ab(ij)}| > 1.96$.

			Respiratory problems during past year			
			Coughing most days during a 3 month period	Bring up phlegm most days during a 3 month period	Wheezing or whistling in chest	Dry cough at night
Lifetime tobacco use	Cigarettes >= 100 times	$\tilde{\theta}_{emp,ij}$ =	2.89	3.31	2.20	2.08
		95% C.I.$_{emp}$ =	(2.02, 4.13)	(2.30, 4.74)	(1.78, 2.73)	(1.33, 3.23)
		$\lvert r_{ab(ij),HA}\rvert$ =	2.14	3.17	0.42	0.07
		$\lvert r_{ab(ij),TR}\rvert$ =	1.08	0.54	0.39	0.98
		$\hat{\theta}_{TR,ij}$ =	2.64	3.11	2.27	2.41
		95% C.I.$_{TR}$ =	(2.00, 3.48)	(2.46, 3.94)	(1.79, 2.87)	(1.57, 3.70)
	Pipe >= 20 times	$\tilde{\theta}_{emp,ij}$ =	2.32	2.56	1.76	1.89
		95% C.I.$_{emp}$ =	(1.64, 3.29)	(1.83, 3.57)	(1.22, 2.54)	(1.09, 3.30)
		$\lvert r_{ab(ij),HA}\rvert$ =	0.72	1.28	1.33	0.43
		$\lvert r_{ab(ij),TR}\rvert$ =	0.91	0.10	0.53	0.20
		$\hat{\theta}_{TR,ij}$ =	2.15	2.53	1.84	1.96
		95% C.I.$_{TR}$ =	(1.52, 3.03)	(1.86, 3.46)	(1.33, 2.56)	(1.27, 3.02)
	Cigars >= 20 times	$\tilde{\theta}_{emp,ij}$ =	1.94	2.09	1.56	1.98
		95% C.I.$_{emp}$ =	(1.48, 2.53)	(1.49, 2.94)	(1.08, 2.27)	(1.28, 3.07)
		$\lvert r_{ab(ij),HA}\rvert$ =	0.94	0.04	2.27	0.31
		$\lvert r_{ab(ij),TR}\rvert$ =	0.42	0.72	0.56	1.30
		$\hat{\theta}_{TR,ij}$ =	1.88	2.22	1.61	1.72
		95% C.I.$_{TR}$ =	(1.40, 2.52)	(1.65, 2.99)	(1.16, 2.24)	(1.12, 2.63)
	Snuff >= 20 times	$\tilde{\theta}_{emp,ij}$ =	1.53	1.79	1.93	2.16
		95% C.I.$_{emp}$ =	(0.85, 2.77)	(1.22, 2.62)	(1.26, 2.95)	(1.17, 3.98)
		$\lvert r_{ab(ij),HA}\rvert$ =	1.25	0.90	0.49	0.07
		$\lvert r_{ab(ij),TR}\rvert$ =	1.14	1.76	1.29	1.03
		$\hat{\theta}_{TR,ij}$ =	1.89	2.23	1.62	1.72
		95% C.I.$_{TR}$ =	(1.38, 2.57)	(1.63, 3.04)	(1.19, 2.20)	(1.11, 2.68)
	Chewing tobacco >= 20 times	$\tilde{\theta}_{emp,ij}$ =	1.68	2.90	1.92	1.77
		95% C.I.$_{emp}$ =	(0.93, 3.05)	(1.81, 4.65)	(1.30, 2.83)	(0.86, 3.66)
		$\lvert r_{ab(ij),HA}\rvert$ =	0.77	1.26	0.49	0.49
		$\lvert r_{ab(ij),TR}\rvert$ =	1.71	1.37	0.51	0.47
		$\hat{\theta}_{TR,ij}$ =	2.11	2.49	1.81	1.93
		95% C.I.$_{TR}$ =	(1.38, 3.24)	(1.59, 3.91)	(1.29, 2.56)	(1.19, 3.13)

Table 5. Estimated type I error rates for SPMI goodness-of-fit statistics. Shaded cells correspond to estimated type I error rates outside of the 95% expected range. For the $(I, J) = (2, 2)$ simulations, $P_{1\bullet(1j)} = 0.2$, $P_{1\bullet(2j)} = 0.3$, $P_{\bullet 1(i1)} = 0.4$, and $P_{\bullet 1(i2)} = 0.5$. For the $(I, J) = (3, 4)$ simulations, $P_{1\bullet(1j)} = 0.3$, $P_{1\bullet(2j)} = 0.4$, $P_{1\bullet(3j)} = 0.5$, $P_{\bullet 1(i1)} = 0.2$, $P_{\bullet 1(i2)} = 0.3$, $P_{\bullet 1(i3)} = 0.4$, and $P_{\bullet 1(i4)} = 0.5$.

$(I, J) = (2, 2)$

Odds ratios within MRCV	Inclusion probabilities	n	X^2_{RS1}	X^2_{RS2}	F 1st	F 2nd
All 2	Large diff.	50	0.055	0.041	0.043	0.029
	Moderate diff.	50	0.044	0.035	0.031	0.025
	Equal	50	0.052	0.041	0.039	0.037
	Large diff.	100	0.060	0.040	0.052	0.020
	Moderate diff.	100	0.062	0.058	0.058	0.036
	Equal	100	0.068	0.064	0.060	0.048
	Large diff.	200	0.054	0.032	0.046	0.022
	Moderate diff.	200	0.066	0.060	0.048	0.038
	Equal	200	0.070	0.062	0.050	0.036
	Large diff.	500	0.038	0.026	0.026	0.014
	Moderate diff.	500	0.076	0.074	0.052	0.050
	Equal	500	0.032	0.032	0.026	0.026
All 25	Large diff.	50	0.106	0.054	0.081	0.044
	Moderate diff.	50	0.065	0.051	0.059	0.035
	Equal	50	0.089	0.052	0.083	0.041
	Large diff.	100	0.110	0.060	0.094	0.026
	Moderate diff.	100	0.110	0.068	0.100	0.038
	Equal	100	0.104	0.054	0.086	0.032
	Large diff.	200	0.106	0.058	0.088	0.034
	Moderate diff.	200	0.102	0.062	0.086	0.038
	Equal	200	0.084	0.048	0.072	0.026
	Large diff.	500	0.090	0.048	0.082	0.024
	Moderate diff.	500	0.090	0.048	0.072	0.042
	Equal	500	0.092	0.062	0.078	0.032

$(I, J) = (3, 4)$

Odds ratios within MRCV	Inclusion probabilities	n	X^2_{RS1}	X^2_{RS2}	F 1st	F 2nd
All 2	Large diff.	100	0.082	0.024	0.082	0.022
	Moderate diff.	100	0.068	0.058	0.066	0.056
	Equal	100	0.062	0.048	0.060	0.042
	Large diff.	200	0.068	0.030	0.064	0.024
	Moderate diff.	200	0.076	0.050	0.070	0.050
	Equal	200	0.074	0.058	0.074	0.054
	Large diff.	500	0.066	0.028	0.066	0.028
	Moderate diff.	500	0.048	0.040	0.046	0.034
	Equal	500	0.052	0.040	0.048	0.038
All 25	Large diff.	100	0.152	0.048	0.146	0.046
	Moderate diff.	100	0.126	0.050	0.124	0.046
	Equal	100	0.131	0.056	0.129	0.056
	Large diff.	200	0.166	0.054	0.162	0.050
	Moderate diff.	200	0.154	0.060	0.150	0.056
	Equal	200	0.120	0.044	0.116	0.040
	Large diff.	500	0.128	0.040	0.124	0.036
	Moderate diff.	500	0.122	0.046	0.118	0.044
	Equal	500	0.110	0.038	0.106	0.036

Table 6. Bias and mean square error for the estimated log-odds ratio. Confidence interval coverage and mean length is included for the true log-odds ratio. The same $P_{ab(ij)}$ values as given in Table 5 are used here.

$(I, J) = (2, 2)$, $n = 200$, all odds ratios are equal to 2

Inclusion probabilities	Bias		MSE		Coverage		Mean length	
	Model	Empirical	Model	Empirical	Model	Empirical	Model	Empirical
Large diff.	0.037	0.046	0.082	0.278	0.944	0.926	1.012	1.790
Moderate diff.	0.009	0.014	0.040	0.138	0.970	0.947	0.805	1.438
Equal	-0.011	-0.005	0.033	0.114	0.956	0.954	0.750	1.348

$(I, J) = (3, 4)$, $n = 200$, all within MRCV odds ratios are equal to 2, $\theta_{1j} = 1$, $\theta_{2j} = 3$, $\theta_{3j} = 5$

Inclusion probabilities	***Y***-item	Bias		MSE		Coverage		Mean length	
		Model	Empirical	Model	Empirical	Model	Empirical	Model	Empirical
Large diff.	1	-0.017	-0.032	0.083	0.267	0.936	0.933	1.017	1.769
	2	0.005	0.014	0.077	0.253	0.928	0.925	0.943	1.706
	3	0.017	0.042	0.073	0.289	0.934	0.933	0.927	1.812
Moderate diff.	1	0.001	-0.007	0.046	0.136	0.946	0.957	0.834	1.436
	2	0.005	0.009	0.039	0.125	0.956	0.952	0.760	1.370
	3	0.003	0.021	0.039	0.153	0.942	0.948	0.747	1.490
Equal	1	-0.002	-0.010	0.040	0.122	0.954	0.955	0.780	1.344
	2	0.009	0.015	0.035	0.108	0.942	0.955	0.711	1.278
	3	0.013	0.030	0.031	0.134	0.954	0.953	0.700	1.395