How Sample Size and Strength of Association Affect the Ability to Detect Group Differences in Cross-Classification Analysis

Raymond Sin-Kwok Wong
University of California-Santa Barbara

February 26, 2003

FIRST DRAFT: PLEASE DO NOT CITE OR
QUOTE WITHOUT PERMISSION FROM AUTHOR

*Paper to be presented at the conference "Social Inequality, Family, and Intergenerational Transfer" of the Research Committee on Social Stratification (RC_28), International Sociological Association in Tokyo, Japan, March 1-3, 2003. Comments are welcome. Please direct all correspondence to Raymond Sin-Kwok Wong, Department of Sociology, University of California, Santa Barbara, CA 93106. Email: wong@soc.ucsb.edu.

Abstract

It is well-known in the statistical literature that when sample size is large, any small departure from the “true” model would yield significant test statistics. This issue is further compounded by the fact that researchers often face the problem of analyzing groups of unequal sizes and as a result our ability to detect group differences may be hampered. Even though this topic has not been investigated systematically, researchers conducting statistical analysis in multi-way cross-classification tables often adopt various *ad hoc* standardization procedures to address this particular problem by inflating and deflating group sizes to comparable values. Using Monte Carlo simulations, this study offers an empirical investigation on the effect of (a) varying sizes across groups, (b) differences in the strength of association between groups, and (c) the ordering of (a) and (b) on our ability to detect differences in association between groups and whether the standardization procedure offers a viable solution. The findings confirm that while it is true that differences in group size and strength of association affect the likelihood-ratio test statistics significantly when the specified models are incorrect, their influence on the “true” model and its over-parameterized counterparts is negligible. In other words, unlike statistical models such as those in covariance structure analysis, the goodness-of-fit statistics in log-linear models are *independent* of sample size when the models specified are “true.” On the other hand, two common problems are associated with the standardization procedure: the problem of underdispersion and the possibility of committing type II error in some occasions. As a result, standardization does not appear to be a viable strategy and researchers should instead rely on conventional nested chi-squared tests for detecting group differences. Finally, the study provides strategies to examine whether the preferred models indeed are the proper ones.

Introduction

Most empirical investigations by social science researchers naturally involve comparisons of different social groups (such as gender, racial/ethnic groups, countries, organizations, and industries) or temporal trend (such as birth cohorts, time series data, and longitudinal data). They include the study of racial and ethnic differences in the relationship between education and occupation, country-specific variation in occupational mobility, and temporal changes in the relationship between education and religiosity, political identification, or attitudes toward abortion. Furthermore, it is not unusual that the sizes of these groups are unequal, due either to population distribution or sampling design. In multiway contingency table analysis, it is generally believed that varying group sizes, especially when the discrepancy is large, may play a pivotal role in influencing our ability to detect group differences or temporal trend. Such belief is further reinforced by the well-known fact in statistical literature that when the sample size is large, any small departure from the "true" model would yield significant test statistics and therefore researchers run considerable risks of identifying trivial differences when in reality these group differences are negligible.[1] As noted by Gelman and Rubin (1995), "it is possible to have so much data that a test will reject every parsimonious model that is proposed" (p. 166).

To date, there is little methodological investigation into this particular subject matter (see Fitzmaurice and Goldthorpe 1997; Fitmaurize, Heath, and Cox 1997; and Wong 1994 for some exceptions). Meanwhile, several 1-*df* tests and other complex statistical models have been developed for cross-classification analysis that partially address this particular problem (Goodman

---

[1]The calculation of the likelihood-ratio chi-squared statistic (L2) is the following:

$$L^2 = -2\ log\ \Lambda = 2\ \sum\sum n_{ij}\ log\ (n_{ij}/\hat{m}_{ij})$$

where Λ is the Wilk statistic, n_{ij} and $\hat{m}_{ij}$ are observed and expected frequencies under a particular model (Agresti 1990)..

and Hout 1998, 2001; Wong 1990, 1995, 2001; Xie 1992; and Yamaguchi 1987). While these statistical models are parsimonious and powerful to detect group differences and therefore offer important tools to empirical researchers, they do not directly address the concern regarding the influence of unequal group size on our ability to detect group differences.

Another strategy to counteract the influence of large and unequal group sizes is to use model selection criteria. The most popular one is the *BIC* (Bayesian Information Criterion) statistic, derived from the Bayesian posteriori test theory (Raftery 1996; but note the critique by Weakliem 1999). Alternatively, other researchers adopt various *ad hoc* standardization procedures to address the problem. They include weighing and standardizing individual tables by (a) using the smallest group as the reference; (b) standardizing them by an arbitrary size (say, 1000), with inflated sizes for some and deflated for others; or (c) rescaling the contribution of individual groups to the likelihood-ratio test statistics by using a standardized index.[2] In some occasions, researchers combine these two strategies together (that is, adopting both standardization and the *BIC* statistic). Of course, there are also works that make *no* correction and assume that varying group sizes have no impact whatsoever in the detection of differences across tables (for instance, Goodman and Hout 2001; Wong 2001).

Given the multitudes of strategies available, empirical researchers may be confused what constitutes the best strategy. With little systematic and empirical investigation of the behavior of these strategies in realistic conditions, researchers and critics often wonder if the adoption of a different strategy may lead to dramatically different conclusions. Furthermore, empirical

[2]In particular, Erikson and Goldthorpe (1993) adopt the following correction in each group: $L^2(S)=((L^2-df)/N)*K+df$ where L^2 is the likelihood-ratio test statistic for a particular group, K is the standardized size, N is the size of the group, and df is the degree of freedom.

researchers would also want to understand the properties of these strategies under various conditions and under what circumstances would one prefer a particular strategy over another.

Generally speaking, we need to consider the following factors in the investigation: (a) the degree of variation in group sizes; (b) the degree of variation in the strength of association across groups; (c) the ordering of (a) and (b); and (d) the number of groups involved in comparison. Owing to the design complexity, the present investigation will consider only two groups and examine only whether the detection of group differences should be based on standardized or unstandardized counts. However, the findings (detailed below) should have broader generalization to analyses with more than two groups and other standardization procedures. In sum, the present study uses Monte Carlo simulations to study the extent to which differences in group size, strength of association and their ordering affect our ability to detect differences and whether statistical analysis should be based on standardized or unstandardized counts.

The present study is partly motivated by the works of Fitzmaurice and Goldthorpe (1997), Fitmaurize, Heath, and Cox (1997), and Wong (1994). However, it differs from them in the following ways. Wong (1994) only considered a limited set of conditions in the Monte Carlo simulations and the major goals of this particular work is to compare how various model selection criteria (*AIC*, *BIC*, nested chi-square tests, and L^2/df ratio) affect our ability to detect group differences. It does not consider the full range of conditions as in this work nor does it consider how the practice of standardization affects the outcome.[3] The study differs from Fitzmaurice and Goldthorpe (1997) and Fitzmaurice, Heath, and Cox (1997), where their major concern is how

---

[3]There is only a close approximation (but not exact relationship) between the simulated data and the estimated models. There is also overdispersion in some of the test statistics (Wong 1994). Both problems are corrected in the present study.

the problem of overdispersion significantly influence our ability to detect group differences.[4] As their works illustrate convincingly, the problem of overdispersion is common in most survey samples, arising from a variety of conditions, design effects (particularly when there is a mix of different sampling schemes), hidden clusters, interviewer effects, and omission of relevant predictors. An important consequence of overdispersion is that it typically favors models that are too complex or rejects any model that is approximately "true," including situations where there is no difference between groups. In sum, the sampling issue and the problem of overdispersion is a persistent and critical subject in empirical research. However, the overdispersion problem will not be directly addressed in the simulation design. It will be included in the discussion section on how to combine findings from Monte Carlo simulations with corrections for overdispersion to examine group differences in association.

Monte Carlo Simulations

Consider the following simple heterogeneous log-multiplicative row and column effects (RC) model (Goodman 1979) for a three-way table with row (R), column (C), and layer (L) variables, with the layer variable representing the grouping variable. The expected frequency for a particular cell, (*i*, *j*, *k*) can be written as the following:

$$log\ m_{ijk} = u + u_i + u_j + u_k + u_{ik} + u_{jk} + \phi_k \mu_i \nu_j \qquad (1)$$

where m_{ijk} is the expected frequency under the model; u is the overall parameter, u_i, u_j, u_k, u_{ik}, and u_{jk} are all one-way or two-way marginal parameters, subject to conventional normalization, ϕ_k is the intrinsic association parameter, and μ_i and ν_j are the estimated row and column scores,

[4]Overdispersion occurs when the data display more variability than is predicted by the variance-mean relationship for the assumed sampling model.

respectively, subject to the following constraints: $\sum \mu_i = \sum \nu_j = 0$ and $\sum \mu_i^2 = \sum \nu_j^2 = 1$. For the sake of simplicity, the simulation study is based on the 4 x 5 x 2 table and the row and column scores are fixed to have the following values: -0.668, -0.198, 0.169, and 0.697 for row scores and -0.610, -0.330, -0.029, 0.329, and 0.640 for column scores. In other words, only ϕ_k are allowed to vary across groups according to the conditions specified below. Depending on the specification of simulated conditions, a total of 50,000 observations are generated for each group and served as the population.[5] Observations are then randomly drawn from each group using the sampling with replacement technique.

Table 1 About Here

By limiting the number of groups to two, we only need to consider three factors in the Monte Carlo simulations: (a) varying size across groups, (b) varying strength of association across groups, and (c) the ordering of variation of (a) and (b). Four different scenarios will be considered for the first two factors: huge, large, moderate, and no difference. To avoid complications arising from sparse cells, the smallest size for any table is 1000. Similarly, the smallest intrinsic association parameter in any group is set to 1. The exact specifications are detailed as the following:

(A) Degree of variation in group size: (a) huge difference (the ratio is 12); (b) large difference (the ratio is 6); (c) small difference (the ratio is 2); and (d) no difference (the ratio is 1);

[5]The master datasets are generated by using the OFFSET command in the statistical software, GLIM (Francis, Green, and Payne 1993).

(B) Degree of variation in the strength of intrinsic association (ϕ) parameter: (a) huge difference (the ratio is 4); (b) large difference (the ratio is 2); (c) small difference (the ratio is 1.5); and (d) no difference (the ratio is 1); and

(C) Ordering of variation in (A) & (B): whether the group with larger size also has greater value in the intrinsic association parameter (ϕ) than the other group and vice versa.

The combination of (A), (B), and (C) yields a total of 25 unique conditions (see Table 1) as the ordering makes no difference for entries in the last row or column (that is, no difference in the strength of association or group size). A total of 100 replicates are then randomly drawn for each condition. In other words, the simulation exercise studies a total of 2,500 three-way tables. The following models are then applied to the tables to test for group differences. They include: (1) conditional independence model (CI), (b) full two-way interaction model (FI), (c) log-linear layer effects model (LL_1), (d) log-multiplicative layer effects model (LL_2), (e) homogeneous uniform association model (UA_1), (f) heterogeneous uniform association model (UA_2), (g) homogeneous log-multiplicative row and column effects model (RC_1), (h) heterogeneous log-multiplicative row and column effects model (RC_2), (i) simple heterogeneous log-multiplicative row and column effects model (RC_3), (j) homogeneous topological model (TOP_1), (k) heterogeneous topological model (TOP_2), and (l) topological model with log-multiplicative layer effect (TOP_3).[6] In particular, RC_1 would be the true model if there is no

[6]The topological model has the following design:

2 2 2 3 3
2 2 2 3 4
3 3 3 1 1
3 3 4 4 1

See Hauser (1978) for details about the utility of topological models in mobility analysis.

difference in strength of association and RC_3 would be the true one when differences occur whereas LL_1, LL_2, and RC_2 can also be regarded as the true but over-parameterized models when differences occur. Finally, the topological models gauge the effect of misspecified models as we usually would not know which one is the true model. The goodness-of-fit statistics will be calculated based on either unstandardized or standardized counts. In the latter case, both groups will be standardized to have 1000 observations.

Results

(a) Effects of Group Size When Difference in Intrinsic Association is Huge

Tables 2-5 report the results of various statistical models to examine the effect of differential group size (huge, large, small, and no difference) when the difference in intrinsic association is huge (that is, ϕ ratio of 4). Judging from the likelihood-ratio test statistics for the conditional independence (CI) and full two-way interaction (FI) models, the results confirm that the effect of sample size can be large for poorly specified or misspecified models. The problem is particularly acute when group size and the strength of association are both huge relative to the other group. For instance, the likelihood-ratio test statistic for CI in condition HH2R is about 4 times larger than HH1R (2796 versus 693). As the discrepancy between the size of the two groups narrows, the ratio of difference in the likelihood-ratio test statistics also narrows, to less than 3 times when the discrepancy in group size is large and only about 1.6 times when the discrepancy is small (see Tables 2, 3, and 4). On the other hand, when groups are standardized to have the same size, the impact of differential group size is under control. The likelihood-ratio test statistics in HH1R, HH2R, LH1R, LH2R, SH1R, and SH2R all display similar values; the range is between 277 and 284. Similar observation can be found for the full two-way interaction model (FI). Thus, when models are misspecified, the standardization procedure eliminates the influence of differential group size. This makes intuitive sense as it is why the procedure is proposed in the first place.

Tables 2-5 About Here

Relative to the full two-way interaction model, the two layer effects model (LL_1 and LL_2) both offer acceptable results when using unstandardized counts. Contrary to expectation, the effect of unequal group size is extremely weak (L^2=10 to 14 with 11 *df*). In the case of the "approximately true" model, the goodness-of-fit statistics for the heterogeneous uniform association (UA_2) model are acceptable at the conventional 0.05 significance level as long as the strength of intrinsic association and differential group size are not concurrently larger than the other group. The L^2 values are 28, 26, and 25 for HH1R, LH1R, SH1R, and NH1R, respectively, with 22 *df*. However, when the strength of intrinsic association and group size within one group are both larger than the other, the likelihood-ratio test statistics are not acceptable when differential group size is either huge (HH2R) or large (LH2R) (56 and 39 points, respectively) but are acceptable in other conditions. While the impact of differential group size is evident here, its impact is much weaker than expected. The range of the test statistics for UA_2 is still quite large (more than 31 chi-squared points apart). The range narrows considerably when standardized counts are used and they are acceptable by conventional level of significance. The same, however, cannot be said to the two over-parameterized models (LL_1 and LL_2), where the problem of underdispersion is evident. That is, the test statistics of the specified models are smaller than their associated degrees of freedom.

When we examine the performance of the true model (the simple heterogeneous RC model, RC_3) and its overparameterized counterpart (heterogeneous RC model, RC_2), we observe that disregarding whether there is any discrepancy in group sizes, the goodness-of-fit statistics for unstandardized counts do not vary across conditions and the L^2/df ratios are very close to 1. This validates our statistical understanding of log-linear modeling that the test statistics and degrees of freedom have the same value when the model is true, a fact that has been widely ignored by empirical researchers who are more concerned with how sample size may affect our ability to detect group differences when there are minor departures from the true model. Again, while the performance of the same models is acceptable when using standardized counts, the problem of underdispersion is self-evident.

Under normal circumstances, researchers would not be able to know *a priori* what the true model is and therefore almost all models estimated are misspecified (in varying degree, of course). This is the reason why we include several topological models (lines 10, 11, 12) in the simulation study. We experiment several different matrices for the topological design through trial-and-error and the one presented here yields the best result. The design matrix does not have the same monotonic relationship between row and column variables as specified in the simulated condition. As expected, the performance of the design matrix is rather poor when the discrepancy in group size is large, especially in HH2R. In general, the results are consistent with the discussion earlier regarding misspecified models: that the standardization procedure is effective in controlling for the impact of differential group size for misspecified models. The test statistics for TOP_1, TOP_2, and TOP_3 are similar in all four tables under standardization. In the case of TOP_3, they are even acceptable at the conventional significance level, though their L^2/df ratios are substantially greater than 1 (between 1.3 to 1.6).

In sum, the above results strongly indicate that our ability to detect group differences in association is not hampered by varying group sizes, particularly when the true model is also specified. Also, whether the calculation is based on standardized or unstandardized counts, the contrasts between models postulating no group difference and group difference are always large and statistically significant. The risk of committing Type I error is extremely small. The standardization procedure is successful in controlling the impact of differential group size on the goodness-of-fit statistics when the models are misspecified. At the same time, the standardization procedure overcorrects the "true" and over-parameterized models, resulting in underdispersion.

(b) Effects of Group Size When Difference in Intrinsic Association is Large

Tables 6-9 report the results of various statistical models to examine the effect of differential group size (huge, large, small, and no difference) when the difference in intrinsic association is large (that is, ϕ ratio of 2). Similar to the case earlier, the impact of differential group size is most noticeable when the difference is huge (a ratio of 12). As expected, the degree of inflated goodness-of-fit statistics is substantially smaller for all misspecified models. For models that are "true" and their over-parameterized counterparts (that is, LL_1, LL_2, RC_2, and RC_3), differential group size again does not affect the test statistics in any meaningful way, the L^2/df ratios are very close to 1.

Tables 6-9 About Here

When we examine the test statistics under standardization, we again observe that the procedure is able to control for the effect of differential group size on the goodness-of-fit statistics when the models are misspecified, with highly comparable test statistics (L_s^2) across the four tables. The larger the degree of model misspecification, the larger the test statistics. On the other hand, the problem of underdispersion persists for the true model and their over-parameterized counterparts. Despite the problem of underdispersion, it should be noted that accurate conclusions about group differences in the association between row and column variables nonetheless can still be reached when comparisons are based on nested chi-squared tests. At the same time, the model postulating full two-way interaction (FI) with no group differences are acceptable under standardization even when there is a huge difference in the size of two groups. The range of the standardized test statistics (L_s^2) is between 16 and 21 with 12 *df*, with a substantial proportion of replicates (over 60 percent in most cases) acceptable at the 0.05 level of significance. This strongly cautions against the use of standardized counts in data analysis. In contrast, the portion is smaller when unstandardized counts are used (less than 10 percent in HL2R and LL2R but over 20 percent in HL1R, LL1R, SL1R, SL2R, and NL1R).

(c) Effects of Group Size When Difference in Intrinsic Association is Small

Tables 10-13 report the results of various statistical models to examine the effect of differential group size (huge, large, small, and no difference) when the difference in intrinsic association is small (that is, ϕ ratio of 1.5). The problem of underdispersion under standardization is most obvious in this particular case. Worse still, many models (FI, UA_1, RC_1, and TOP_1) that postulate no group difference even achieve acceptable goodness-of-fit statistics (L_s^2) and the L_s^2/df ratios are close to or smaller than 1 in some occasions. If researchers based on their analyses solely on standardized counts, they would have concluded that there are no group differences in association and therefore running the risk of committing type II error (that is, incorrectly accepting a wrong model or not rejecting a false hypothesis). Furthermore, one cannot rely on the nested chi-squared tests to gauge evidence of group differences in association.

Clearly, standardization is not the preferred strategy when group differences in association are small. The procedure overcorrects and becomes problematic when the statistical power of the tests are relatively low.

Tables 10-13 About Here

On the other hand, the problem is much less obvious when unstandardized counts are used. While it is true that the goodness-of-fit of some misspecified models (that is, models that postulate no group differences in association) are acceptable at the conventional level (for example, as high as 84 percent for the full two-way interaction model in LS1R), all of them have the L^2/df ratios substantially greater than 1. Not surprisingly, for the "true" model (RC_3) and its over-parameterized counterparts (LL_1, LL_2, and RC_2), the ratio of L^2 relative to its df continues to be very close to 1. On average, the nested chi-squared tests appear to work moderately well.

(d) Effects of Group Size When Difference in Intrinsic Association is None

Tables 14-17 report the results of various statistical models to examine the effect of differential group size (huge, large, small, and no difference) when there is no difference in intrinsic association (that is, ϕ ratio of 1). Again, the problem of underdispersion is evident when adjustments are made to make the two groups comparable. Fortunately, unlike the previous case, decisions about group differences do not seem to hamper our decision about group differences. With the exception of the conditional independence model, the L^2/df ratios for all other models are substantially less than 1.

Tables 14-17 About Here

The pattern is rather similar when raw counts are used. With the exception of the conditional independence (CI) and the two uniform association models (UA_1 and UA_2), the L^2/df ratios for all other models are very close to 1. Correct decisions about no group differences can be obtained from the nested chi-squared tests as none of them are statistically significant at the 0.05 level. In sum, the findings here indicate that when there is no group difference in association, the performance of models based on standardized counts are just as adequate as those based on raw counts. Analysis based on raw counts are thus preferred because there is no problem of underdispersion, created artificially by the weighting procedure itself.

Using Nested Chi-Squared Tests to Detect Group Differences

The above findings cast doubts against the use of any weighting procedure to correct for differential group size in log-linear modeling. At the same time, they also point to the utility of nested chi-squared tests to detect group differences (Weakliem 1999). To illustrate further the performance of the nested chi-squared tests to study group differences, Table 18 summarizes the contrasts between several homogeneous and heterogeneous models. A total of seven nested tests are conducted to check the number of times one can reject the model of no group difference in association, using the 0.05 level of statistical significance as the cutoff. They include tests between FI and LL_1 (Δ_1); FI and LL_2 (Δ_2); UA_1 and UA_2 (Δ_3); RC_1 and RC_2 (Δ_4); RC_1 and RC_3 (Δ_5); TOP_1 and TOP_2 (Δ_6); and TOP_1 and TOP_3 (Δ_7) for both unstandardized and standardized counts. Each entry represents the number of times that the null hypothesis of no group difference can be rejected.

Table 18 About Here

When the underlying difference in association is huge between groups, the nested chi-square difference tests are able to reject the null hypothesis without difficulty, disregarding whether the specified models conform to the underlying true model (lines 1 through 7). The probability of rejecting the null hypothesis is identical for both raw and standardized counts. However, as the difference in association between groups declines, not all tests perform equally well. As expected, the contrasts between RC_1 and RC_3 (Δ_5) provides the best result and is therefore used as the benchmark for comparison. When the difference in ϕ ratio is large, we can reject more than 96 out of 100 times in Δ_5 whenever the two groups are of different sizes, no matter how small that discrepancy is. The performance deteriorates slightly to only 89 percent when the two groups have equal size. The performance of other contrasts is highly comparable to Δ_5, particularly those involving 1-*df* tests. Contrasts based on standardized counts also yield comparable results but they are clearly somewhat inferior than their unstandardized counterparts.

The undesirability of using standardized counts is most obvious when the difference in association between groups is small. By comparing the values for Δ_5 and Δ_{5s}, the performance of the former clearly outpaced that of the latter, especially in the case for HS2R, LS2R, and SS2R, that is, when the size and strength of association of one group are both higher than the other. In these cases, the number of times that one can reject the null hypothesis can be three times larger when using raw counts as opposed to standardized counts. The standardization procedure clearly overcorrects and as a result the risk of committing type II error increases. Admittedly, the performance of these nested tests are not high (between 43 and 75 percent) even when raw counts are used in the analysis. Nonetheless, their performance is far more superior than any other alternatives.

The only condition that standardized counts provide better results is when there is no difference in association across groups. Entries in the last four rows consistently show that using standardized counts yield the lowest number of times that one would reject the "true" null hypothesis of no difference in association. The performance of nested tests based on raw counts is highly comparable and the risks of committing type I error for most contrasts are acceptable. In sum, the above results clearly shows that the standardization procedure is a conservative strategy and has a biased tendency against group differences in association. Since the performance of nested tests based on raw counts perform consistently well in all scenarios, there is no need to make adjustment for differential group size and one should continue to rely on the conventional nested chi-squared tests to test for differences.

Suggestions for Modeling Strategies

The Monte Carlo experiment provides the following important findings that can be used to inform modeling strategies. First, there is *no* impact of sample size (differential group size, in particular) on our ability to detect group differences when the specified model is true. Second, there is *no* impact in the differential strength of association on our ability to detect group differences when the specified model is true. In both cases, the value of the goodness-of-fit statistic relative to its associated degrees of freedom is always close to 1. Third, only when models that depart severely from the true model would one find the influence of sample size and strength of association on the goodness-of-fit statistics. In other words, unlike statistical models in covariance structure analysis, the calculation of goodness-of-fit statistics in *independent* of sample size and strength of association, a fact that has been consistently ignored by applied researchers who tend to focus on misspecified models instead. An important corollary of this finding is that using the proportion reduction in L^2 or commonly known as the normed fit index (NFI) in covariance structure analysis to justify a particular non-fitting model is problematic, especially when the conditional independence model, often expected to be incorrect in the first place, is chosen as the baseline model. In other words, a 96 or 98 proportion reduction in goodness-of-fit statistics does not guarantee the model under consideration is indeed a good one. Fourth, the performance of over-parameterized models rivals that of the true model and therefore can be used as competing models during the modeling stage.

In statistical modeling, scientific accuracy and scientific parsimony are both important and they are not necessarily tradeoffs. For most practitioners, the major reason why they opt for correction or adopt other *ad hoc* strategies is because none of the models fit the data well. The Monte Carlo experiment clearly demonstrate that the reason why they do not fit well is not due to their huge size or underlying association, but rather because the models are "misspecified." The reasons for misspecification are many but the most obvious one, of course, is that the models are too simple to handle the complex association pattern governing the data. In other words, we should work harder to search for more complex models instead. A number of them are already widely available in the literature and the list below does not meant to be exhaustive. The only caveat is that the choice of these alternative models should not be based on methodological prowess, but theoretical import and substantive interpretation. In addition to the models discussed earlier, other complex models include: (a) hybrid models that combine vertical and non-vertical effects (Stier and Grusky 1990; Wong 1992); (b) Goodman's U+RC, R+RC, C+RC model (Goodman 1986); (d) Goodman's multidimensional association models for groups, RC(M)-G models (Becker 1989; Becker and Clogg 1989); (d) log-trilinear multidimensional scaled association models that use either the PARAFAC/CANDECOMP or 3-mode decomposition method (Wong 2001); and (e) the modified regression-type models (Goodman and Hout 1998, 2001). All of them have been successfully applied to tables that yield new and interesting understanding of the underlying association between tables.

Two modeling strategies can be adopted. The first one uses bootstrapping and resampling techniques and the second one use arbitrary scaling and rescaling methods. In the case of bootstrapping or resampling, the following steps can be taken: (a) divide the tables into two or multiple tables randomly: If the sample size is small, use sampling with replacement to generate multiple samples; (b) estimate a wide variety of models and use either L^2 or L^2/df as the criterion for choosing models to a particular sample; (c) apply the same set of models to other generated samples: if the same preferred models fit consistently well, they are potential "true" models. Use Bayesian or other model selection criteria to guide final selection. Often times, it is appropriate to report all competing models so that others can judge whether alternative interpretation is possible.

The second method uses arbitrary scaling and rescaling methods as the simulation study finds that any scaling of the size of one group relative to another bears no consequence to the true model or its over-parameterized counterparts. The following steps can be taken: (a) assign arbitrary weights to each table and estimate a wide variety of models and use either L^2 or L^2/df as the criterion for choosing models; (b) repeat step (a) by using other arbitrary weights and estimate the same set of models for at least three or more times; and (c) compare the performance of all competing models and use Bayesian or other model selection criteria to guide final selection. All competing alternative models can be further evaluated by using non-nested models tests (Weakliem 1992).

Both strategies require more than one set of model estimations. This is necessary in order to avoid the problem of capitalizing on chance and overfitting in a particular situaton. Note that the suggestions assume the existence of sufficient cases for each cell. Sparse cells can pose a major problem but the problem may be less severe in certain parameteric models and in some occasions, their location in the tables may or may not be consequential.

The Problem of OverDispersion

The discussion so far ignores the problem of overdispersion, which is extremely common in survey data. As McCullagh and Nelder (1989:125) remark, it is generally wise to assume the presence of overdispersion unless the data or prior information suggest otherwise. Since the problem of overdispersion is often related to sampling, it should be corrected before statistical modeling. Fitmaurice, Heath, and Cox (1997) and Fitmaurice and Goldthorpe (1997) offer rather simple solution to this particular problem. However, once the correction is made, we should continue to rely on L^2, L^2/df ratio of 1, and nested chi-squared tests in model comparison and avoid any further correction on sample size and alike.

Conclusion

The concern about the impact of unequal sample sizes and underlying strength of association from different tables in cross-classification analysis is a *misplaced* one. Not only that there is no need to adjust the test statistics through standardization, the procedure may inadvertently create the problem of underdispersion, thereby giving researchers a false security that the no difference model actually fits well when group differences occur. Since differences in group size and strength of association play no role in the likelihood-ratio test statistics when the specified model is true, researchers should continue to rely on nominal test statistics and nested chi-squared tests for group differences. The use of overparameterized and complex models should also be encouraged in empirical investigations. They would provide researchers better alternatives to choose rather than relying simple but poorly specified models. Finally, the two modeling strategies should be applied routinely in future empirical applications as they would help researchers to gain confidence that their final models approximate the underlying "true" model. It is only when researchers can entertain several (but not one) competing models, then we are likely to achieve the twin goals of science: accuracy and parsimony.

Bibliography

Anderson, Carolyn J. 1996. "The Analysis of Three-Way Contingency Tables by Three-Mode Association Models." *Psychometrika* 61:465-43.

Becker, Mark P. 1989. "Models for the Analysis of Association in Multiway Contingency Tables." *Journal of the American Statistical Association* 84:1014-19.

Becker, Mark P. and Clifford C. Clogg. 1989. "Analysis of Sets of Two-Way Contingency Tables Using Association Models." *Journal of the American Statistical Association* 84:142-51.

Cox, David R. 1983. "Some Remarks on Overdispersion." *Biometrika* 70:269-74.

Erikson, Robert and John H. Goldthorpe. 1993. *The Constant Flux: A Study of Class Mobility in Industrial Societies.* Oxford: Clarendon Press.

Fitzmaurice, Garrett M., Anthony F. Heath, and David R. Cox. 1997. "Detecting Overdispersion in Large Scale Surveys: Application to a Study of Education and Social Class in Britain." *Royal Statistical Society. Journal. Series C: Applied Statistics* 46:415-426.

Fitzmaurice, G. M and John H. Goldthorpe. 1997. "Adjusting for Overdispersion in an Analysis of Comparative Social Mobility." *Sociological Methods & Research* 25:267-83.

Francis, B., M. Green, and C. Payne. 1993. *The GLIM System: Release 4 Manual*. Oxford, U.K.: Clarendon.

Goodman, Leo A. 1979. "Simple Models for the Analysis of Association in Cross-Classifications Having Ordered Categories." *Biometrika* 68:347-55.

----------. 1986. "Some Useful Extensions of the Usual Correspondence Analysis Approach and the Usual Log-Linear Models Approach in the Analysis of Contingency Tables (with Discussion)." *International Statistical Review* 54:243-70.

Goodman, Leo A. and Michael Hout. 1998. "Statistical Methods and Graphical Displays for Analyzing How the Association Between Two Qualitative Variables Differ Among Countries, Among Groups or Over Time: A Modified Regression-Type Approach." Pp. 175-230 in *Sociological Methodology*, Vol. 28, edited by Adrian Raftery. Washington, DC: American Sociological Association.

----------. 2001. "Statistical Methods and Graphical Displays for Analyzing How the Association Between Two Qualitative Variables Differ Among Counries, Among Groups or Over Time. Part II: Some Exploratory Techniques, Simple Models, and Simple Examples." Pp. 189-221 in *Sociological Methodology*, Vol. 31, edited by Mark Becker. Washington, DC: American Sociological Association.

Hauser, Robert M. 1978. "A Structural Model of the Mobility Table." *Social Forces* 56:919-53.

McCullagh, P. and J. A. Nelder. 1989. *Generalized Linear Models*. 2nd edition. London: Chapman & Hall.

Raftery, Adrian E. 1996. "Bayesian Model Selection in Social Research." Pp. 111-163 in *Sociological Methodology*, Vol, 25, edited by Peter V. Marsden. Washington, DC: American Sociological Association.

Stier, Haya and David B. Grusky. 1990. "An Overlapping Persistence Model of Career Mobility." *American Sociological Review* 55:736-56.

Vermunt, Jeroen. 1997. *LEM: A General Program for the Analysis of Categorical Data*. Version 1.0. Tilburg University, Netherlands.

Weakliem, David. 1992. "Comparing Non-Nested Models for Contingency Tables." Pp.147-178 in *Sociological Methodology, Vol. 22*, edited by Peter V. Marsden. Washington, DC: American Sociology Association.

----------. 1999. "A Critique of the Bayesian Information Criterion for Model Selection." *Sociological Methods & Research* 27:359-97.

Wong, Raymond Sin-Kwok. 1990. "Understanding Cross-National Variation in Occupational Mobility." *American Sociological Review* 55:560-573.

----------. 1992. "Vertical and Non-Vertical Effects in Class Mobility: Cross-National Variations." *American Sociological Review* 57:396-410.

----------. 1994. "Model Selection Strategies and the Use of Association Models to Detect Group Differences." *Sociological Methods & Research* 22:460-91.

----------. 2001. "Multidimensional Association Models: A Multilinear Approach." *Sociological Methods & Research* 30:197-240.

Xie, Yu. 1992. "The Log-Multiplicative Layer Models for Comparing Mobility Tables." *American Sociological Review* 57:380-95.

Yamaguchi, Kazuo. 1987. "Models for Comparing Mobility Tables: Towards Parsimony and Substance." *American Sociological Review* 52:482-94.

Table 1
Simulated Conditions to Study the Effects of Sample Size and Strength of Association on the Detection of Group Differences

Sample Size (N)	Intrinsic Association (ϕ) Huge Difference (ϕ ratio=4)	Large Difference (ϕ ratio=2)	Small Difference (ϕ ratio=1.5)	No Difference (ϕ ratio=1)
Huge Difference (size ratio=12)	1, 2	3, 4	5, 6	7
Large Difference (size ratio=6)	8, 9	10, 11	12, 13	14
Small Difference (size ratio=2)	15, 16	17, 18	19, 20	21
No Difference (size ratio=1)	22	23	24	25

Note: 100 simulated tables are constructed for each simulated condition. Entries with two conditions indicate that the ordering of group size and strength of association may be consequential.

Table 2
Condition 1: Huge Difference in Association, Huge Difference in Group Size

Model Description	*df*	L^2	*p*	L_s^2	*p*
(a) HH1R: N_1=1,000, ϕ_1=4.0; N_2=12,000, ϕ_2=1.0					
1. Conditional Independence Model (CI)	24	693.1	0.00	280.8	0.00
2. Full Two-Way Interaction Model (FI)	12	126.9	0.00	82.7	0.00
3. Log-Linear Layer Effect Model (LL_1)	11	13.7	0.89	8.9	0.99
4. Log-Multiplicative Layer Effect Model (LL_2)	11	11.6	0.96	2.6	1.00
5. Homogeneous UA Model (UA_1)	23	139.8	0.00	93.3	0.00
6. Heterogeneous UA Model (UA_2)	22	28.3	0.77	16.2	1.00
7. Homogeneous RC Model (RC_1)	18	133.8	0.00	89.6	0.00
8. Heterogeneous RC Model (RC_2)	12	12.4	0.95	7.2	1.00
9. Simple Heterogeneous RC Model (RC_3)	17	17.3	0.94	8.2	1.00
10. Homogeneous Topological Model (TOP_1)	20	157.4	0.00	102.1	0.00
11. Heterogeneous Topological Model (TOP_2)	16	46.3	0.03	24.3	0.69
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	50.0	0.02	25.2	0.76
(b) HH2R: N_1=12,000, ϕ_1=4.0; N_2=1,000, ϕ_2=1.0					
1. Conditional Independence Model (CI)	24	2796.1	0.00	276.9	0.00
2. Full Two-Way Interaction Model (FI)	12	209.0	0.00	77.0	0.00
3. Log-Linear Layer Effect Model (LL_1)	11	12.8	0.84	5.6	1.00
4. Log-Multiplicative Layer Effect Model (LL_2)	11	10.6	0.95	9.2	0.99
5. Homogeneous UA Model (UA_1)	23	247.2	0.00	90.2	0.00
6. Heterogeneous UA Model (UA_2)	22	56.1	0.04	14.6	1.00
7. Homogeneous RC Model (RC_1)	18	336.8	0.00	84.6	0.00
8. Heterogeneous RC Model (RC_2)	12	11.1	0.99	5.9	1.00
9. Simple Heterogeneous RC Model (RC_3)	17	16.3	0.96	10.5	1.00
10. Homogeneous Topological Model (TOP_1)	20	392.4	0.00	98.4	0.00
11. Heterogeneous Topological Model (TOP_2)	16	188.7	0.00	23.9	0.74
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	191.7	0.00	26.5	0.77

Table 3
Condition 2: Huge Difference in Association, Large Difference in Group Size

Model Description	*df*	L^2	*p*	L_s^2	*p*
(a) LH1R: N_1=1,000, ϕ_1=4.0; N_2=6,000, ϕ_2=1.0					
1. Conditional Independence Model (CI)	24	472.3	0.00	276.5	0.00
2. Full Two-Way Interaction Model (FI)	12	115.8	0.00	79.2	0.00
3. Log-Linear Layer Effect Model (LL_1)	11	11.9	0.95	8.2	1.00
4. Log-Multiplicative Layer Effect Model (LL_2)	11	11.1	0.95	3.3	1.00
5. Homogeneous UA Model (UA_1)	23	129.1	0.00	89.3	0.00
6. Heterogeneous UA Model (UA_2)	22	25.7	0.88	15.2	1.00
7. Homogeneous RC Model (RC_1)	18	122.8	0.00	85.2	0.00
8. Heterogeneous RC Model (RC_2)	12	11.5	0.96	6.7	1.00
9. Simple Heterogeneous RC Model (RC_3)	17	16.2	0.97	8.0	1.00
10. Homogeneous Topological Model (TOP_1)	20	139.7	0.00	98.1	0.00
11. Heterogeneous Topological Model (TOP_2)	16	37.1	0.12	23.8	0.67
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	40.3	0.12	24.8	0.78
(b) LH2R: N_1=6,000, ϕ_1=4.0; N_2=1,000, ϕ_2=1.0					
1. Conditional Independence Model (CI)	24	1428.4	0.00	277.5	0.00
2. Full Two-Way Interaction Model (FI)	12	178.2	0.00	78.6	0.00
3. Log-Linear Layer Effect Model (LL_1)	11	12.2	0.92	5.8	1.00
4. Log-Multiplicative Layer Effect Model (LL_2)	11	10.3	0.96	8.9	0.99
5. Homogeneous UA Model (UA_1)	23	205.1	0.00	90.4	0.00
6. Heterogeneous UA Model (UA_2)	22	38.6	0.35	15.3	1.00
7. Homogeneous RC Model (RC_1)	18	195.3	0.00	85.8	0.00
8. Heterogeneous RC Model (RC_2)	12	11.9	0.92	6.6	1.00
9. Simple Heterogeneous RC Model (RC_3)	17	16.6	0.94	10.8	1.00
10. Homogeneous Topological Model (TOP_1)	20	275.1	0.00	98.3	0.00
11. Heterogeneous Topological Model (TOP_2)	16	101.6	0.00	24.1	0.75
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	104.4	0.00	26.6	0.82

Table 4
Condition 3: Huge Difference in Association, Small Difference in Group Size

Model Description	*df*	L^2	*p*	L^2_s	*p*
(a) SH1R: N_1=1,000, ϕ_1=4.0; N_2=2,000, ϕ_2=1.0					
1. Conditional Independence Model (CI)	24	322.6	0.00	280.8	0.00
2. Full Two-Way Interaction Model (FI)	12	101.4	0.00	82.6	0.00
3. Log-Linear Layer Effect Model (LL_1)	11	11.8	0.95	9.6	0.99
4. Log-Multiplicative Layer Effect Model (LL_2)	11	10.9	1.00	6.3	1.00
5. Homogeneous UA Model (UA_1)	23	116.1	0.00	95.0	0.00
6. Heterogeneous UA Model (UA_2)	22	24.8	0.91	19.1	0.98
7. Homogeneous RC Model (RC_1)	18	110.3	0.00	90.3	0.00
8. Heterogeneous RC Model (RC_2)	12	11.9	0.97	8.9	1.00
9. Simple Heterogeneous RC Model (RC_3)	17	16.9	1.00	11.7	1.00
10. Homogeneous Topological Model (TOP_1)	20	124.1	0.00	103.6	0.00
11. Heterogeneous Topological Model (TOP_2)	16	33.0	0.32	27.5	0.47
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	36.1	0.28	29.2	0.58
(b) SH2R: N_1=2,000, ϕ_1=4.0; N_2=1,000, ϕ_2=1.0					
1. Conditional Independence Model (CI)	24	519.0	0.00	284.2	0.00
2. Full Two-Way Interaction Model (FI)	12	120.0	0.00	80.6	0.00
3. Log-Linear Layer Effect Model (LL_1)	11	12.5	0.89	8.7	0.99
4. Log-Multiplicative Layer Effect Model (LL_2)	11	10.7	0.93	9.8	0.96
5. Homogeneous UA Model (UA_1)	23	141.4	0.00	95.2	0.00
6. Heterogeneous UA Model (UA_2)	22	29.1	0.71	20.1	0.96
7. Homogeneous RC Model (RC_1)	18	133.7	0.00	89.3	0.00
8. Heterogeneous RC Model (RC_2)	12	12.0	0.91	8.9	0.98
9. Simple Heterogeneous RC Model (RC_3)	17	16.7	0.94	13.4	0.97
10. Homogeneous Topological Model (TOP_1)	20	152.3	0.00	102.1	0.00
11. Heterogeneous Topological Model (TOP_2)	16	44.7	0.06	26.6	0.56
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	48.2	0.09	29.8	0.52

Table 5
Condition 4: Huge Difference in Association, No Difference in Group Size

Model Description	*df*	L^2	*p*
(a) NH1R: N_1=1,000, ϕ_1=4.0; N_2=1,000, ϕ_2=1.0			
1. Conditional Independence Model (CI)	24	293.6	0.00
2. Full Two-Way Interaction Model (FI)	12	88.7	0.00
3. Log-Linear Layer Effect Model (LL_1)	11	12.3	0.88
4. Log-Multiplicative Layer Effect Model (LL_2)	11	11.2	0.96
5. Homogeneous UA Model (UA_1)	23	105.3	0.00
6. Heterogeneous UA Model (UA_2)	22	26.4	0.81
7. Homogeneous RC Model (RC_1)	18	105.3	0.00
8. Heterogeneous RC Model (RC_2)	12	12.6	0.97
9. Simple Heterogeneous RC Model (RC_3)	17	17.5	0.94
10. Homogeneous Topological Model (TOP_1)	20	111.5	0.00
11. Heterogeneous Topological Model (TOP_2)	16	31.9	0.31
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	34.9	0.37

Table 6

Condition 5: Large Difference in Association, Huge Difference in Group Size

Model Description	*df*	L^2	*p*	L^2_s	*p*
(a) HL1R: N_1=1,000, ϕ_1=2.0; N_2=12,000, ϕ_2=1.0					
1. Conditional Independence Model (CI)	24	524.2	0.00	120.9	0.00
2. Full Two-Way Interaction Model (FI)	12	28.3	0.20	17.9	0.67
3. Log-Linear Layer Effect Model (LL_1)	11	10.9	0.94	6.5	1.00
4. Log-Multiplicative Layer Effect Model (LL_2)	11	10.5	0.94	3.3	1.00
5. Homogeneous UA Model (UA_1)	23	42.0	0.33	23.7	0.92
6. Heterogeneous UA Model (UA_2)	22	25.8	0.89	12.9	1.00
7. Homogeneous RC Model (RC_1)	18	34.6	0.30	21.1	0.85
8. Heterogeneous RC Model (RC_2)	12	12.1	0.93	6.4	1.00
9. Simple Heterogeneous RC Model (RC_3)	17	16.7	0.97	8.0	1.00
10. Homogeneous Topological Model (TOP_1)	20	53.5	0.03	26.1	0.75
11. Heterogeneous Topological Model (TOP_2)	16	34.5	0.27	13.4	0.96
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	37.3	0.25	14.3	0.98
(b) HL2R: N_1=12,000, ϕ_1=2.0; N_2=1,000, ϕ_2=1.0					
1. Conditional Independence Model (CI)	24	928.6	0.00	122.2	0.00
2. Full Two-Way Interaction Model (FI)	12	38.3	0.06	15.9	0.87
3. Log-Linear Layer Effect Model (LL_1)	11	11.5	0.90	5.6	1.00
4. Log-Multiplicative Layer Effect Model (LL_2)	11	11.1	0.94	8.2	0.99
5. Homogeneous UA Model (UA_1)	23	56.9	0.02	23.0	0.96
6. Heterogeneous UA Model (UA_2)	22	33.3	0.54	13.1	1.00
7. Homogeneous RC Model (RC_1)	18	44.7	0.10	19.7	0.91
8. Heterogeneous RC Model (RC_2)	12	12.1	0.93	6.8	0.99
9. Simple Heterogeneous RC Model (RC_3)	17	17.0	0.92	10.4	1.00
10. Homogeneous Topological Model (TOP_1)	20	91.0	0.00	25.6	0.81
11. Heterogeneous Topological Model (TOP_2)	16	62.9	0.00	14.5	0.96
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	65.4	0.00	16.4	0.97

Table 7
Condition 6: Large Difference in Association, Large Difference in Group Size

Model Description	*df*	L^2	*p*	L_s^2	*p*
(a) LL1R: N_1=1,000, ϕ_1=2.0; N_2=6,000, ϕ_2=1.0					
1. Conditional Independence Model (CI)	24	315.0	0.00	123.1	0.00
2. Full Two-Way Interaction Model (FI)	12	27.5	0.23	18.3	0.75
3. Log-Linear Layer Effect Model (LL_1)	11	11.3	0.91	7.2	1.00
4. Log-Multiplicative Layer Effect Model (LL_2)	11	11.2	0.94	4.4	1.00
5. Homogeneous UA Model (UA_1)	23	33.4	0.35	21.6	0.95
6. Heterogeneous UA Model (UA_2)	22	24.0	0.89	13.7	1.00
7. Homogeneous RC Model (RC_1)	18	33.4	0.30	21.6	0.89
8. Heterogeneous RC Model (RC_2)	12	11.9	0.95	6.9	1.00
9. Simple Heterogeneous RC Model (RC_3)	17	16.9	0.95	8.8	1.00
10. Homogeneous Topological Model (TOP_1)	20	44.8	0.08	26.7	0.80
11. Heterogeneous Topological Model (TOP_2)	16	27.0	0.53	14.3	0.96
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	29.7	0.57	15.4	0.97
(b) LL2R: N_1=6,000, ϕ_1=2.0; N_2=1,000, ϕ_2=1.0					
1. Conditional Independence Model (CI)	24	500.1	0.00	124.0	0.00
2. Full Two-Way Interaction Model (FI)	12	34.7	0.07	16.4	0.80
3. Log-Linear Layer Effect Model (LL_1)	11	11.1	0.93	5.9	1.00
4. Log-Multiplicative Layer Effect Model (LL_2)	11	10.8	0.94	8.2	1.00
5. Homogeneous UA Model (UA_1)	23	48.7	0.11	23.8	0.94
6. Heterogeneous UA Model (UA_2)	22	27.0	0.83	13.7	1.00
7. Homogeneous RC Model (RC_1)	18	40.6	0.15	19.8	0.96
8. Heterogeneous RC Model (RC_2)	12	11.0	1.00	6.2	1.00
9. Simple Heterogeneous RC Model (RC_3)	17	16.4	0.98	10.4	1.00
10. Homogeneous Topological Model (TOP_1)	20	64.0	0.01	25.2	0.82
11. Heterogeneous Topological Model (TOP_2)	16	38.2	0.15	13.3	1.00
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	41.6	0.17	16.0	1.00

Table 8
Condition 7: Large Difference in Association, Small Difference in Group Size

Model Description	*df*	L^2	*p*	L_s^2	*p*
(a) SL1R: N_1=1,000, ϕ_1=2.0; N_2=2,000, ϕ_2=1.0					
1. Conditional Independence Model (CI)	24	171.1	0.00	129.1	0.00
2. Full Two-Way Interaction Model (FI)	12	26.0	0.35	20.9	0.61
3. Log-Linear Layer Effect Model (LL_1)	11	11.3	0.95	8.8	1.00
4. Log-Multiplicative Layer Effect Model (LL_2)	11	10.8	0.95	6.7	1.00
5. Homogeneous UA Model (UA_1)	23	37.4	0.48	29.3	0.75
6. Heterogeneous UA Model (UA_2)	22	23.3	0.91	17.7	0.98
7. Homogeneous RC Model (RC_1)	18	32.1	0.44	25.5	0.70
8. Heterogeneous RC Model (RC_2)	12	11.7	0.97	8.8	0.99
9. Simple Heterogeneous RC Model (RC_3)	17	16.9	0.95	12.1	0.99
10. Homogeneous Topological Model (TOP_1)	20	39.3	0.27	31.0	0.56
11. Heterogeneous Topological Model (TOP_2)	16	22.0	0.74	16.8	0.91
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	25.1	0.73	18.7	0.96
(b) SL2R: N_1=2,000, ϕ_1=2.0; N_2=1,000, ϕ_2=1.0					
1. Conditional Independence Model (CI)	24	208.2	0.00	127.4	0.00
2. Full Two-Way Interaction Model (FI)	12	29.0	0.21	20.1	0.56
3. Log-Linear Layer Effect Model (LL_1)	11	11.1	0.95	8.0	1.00
4. Log-Multiplicative Layer Effect Model (LL_2)	11	10.8	0.98	9.4	0.99
5. Homogeneous UA Model (UA_1)	23	40.3	0.34	28.6	0.82
6. Heterogeneous UA Model (UA_2)	22	23.2	0.95	17.0	1.00
7. Homogeneous RC Model (RC_1)	18	34.9	0.27	24.5	0.74
8. Heterogeneous RC Model (RC_2)	12	11.8	0.95	8.8	0.99
9. Simple Heterogeneous RC Model (RC_3)	17	16.5	0.98	12.9	1.00
10. Homogeneous Topological Model (TOP_1)	20	44.6	0.08	30.4	0.64
11. Heterogeneous Topological Model (TOP_2)	16	24.6	0.62	16.8	0.96
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	27.4	0.69	19.2	0.98

Table 9
Condition 8: Large Difference in Association, No Difference in Group Size

Model Description	*df*	L^2	*p*
(a) NL1R: N_1=1,000, ϕ_1=2.0; N_2=1,000, ϕ_2=1.0			
1. Conditional Independence Model (CI)	24	135.9	0.00
2. Full Two-Way Interaction Model (FI)	12	21.9	0.52
3. Log-Linear Layer Effect Model (LL_1)	11	11.5	0.96
4. Log-Multiplicative Layer Effect Model (LL_2)	11	11.4	0.91
5. Homogeneous UA Model (UA_1)	23	33.3	0.61
6. Heterogeneous UA Model (UA_2)	22	23.4	0.92
7. Homogeneous RC Model (RC_1)	18	28.4	0.55
8. Heterogeneous RC Model (RC_2)	12	12.5	0.94
9. Simple Heterogeneous RC Model (RC_3)	17	17.7	0.94
10. Homogeneous Topological Model (TOP_1)	20	34.2	0.42
11. Heterogeneous Topological Model (TOP_2)	16	21.3	0.77
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	24.2	0.78

Table 10
Condition 9: Small Difference in Association, Huge Difference in Group Size

Model Description	*df*	L^2	*p*	L_s^2	*p*
(a) HS1R: N_1=1,000, ϕ_1=1.5; N_2=12,000, ϕ_2=1.0					
1. Conditional Independence Model (CI)	24	506.9	0.00	93.3	0.00
2. Full Two-Way Interaction Model (FI)	12	16.7	0.75	10.3	1.00
3. Log-Linear Layer Effect Model (LL_1)	11	11.4	0.95	6.9	0.99
4. Log-Multiplicative Layer Effect Model (LL_2)	11	11.3	0.95	4.7	1.00
5. Homogeneous UA Model (UA_1)	23	31.3	0.70	16.2	1.00
6. Heterogeneous UA Model (UA_2)	22	26.6	0.77	13.2	1.00
7. Homogeneous RC Model (RC_1)	18	23.0	0.77	13.5	1.00
8. Heterogeneous RC Model (RC_2)	12	12.4	0.92	6.7	1.00
9. Simple Heterogeneous RC Model (RC_3)	17	17.6	0.92	9.1	1.00
10. Homogeneous Topological Model (TOP_1)	20	43.1	0.18	18.1	0.98
11. Heterogeneous Topological Model (TOP_2)	16	35.7	0.18	13.3	1.00
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	38.5	0.26	14.6	0.99
(b) HS2R: N_1=12,000, ϕ_1=1.5; N_2=1,000, ϕ_2=1.0					
1. Conditional Independence Model (CI)	24	575.4	0.00	91.2	0.00
2. Full Two-Way Interaction Model (FI)	12	19.3	0.63	8.8	1.00
3. Log-Linear Layer Effect Model (LL_1)	11	10.9	0.92	5.5	1.00
4. Log-Multiplicative Layer Effect Model (LL_2)	11	10.8	0.92	6.9	1.00
5. Homogeneous UA Model (UA_1)	23	34.5	0.55	15.3	1.00
6. Heterogeneous UA Model (UA_2)	22	27.5	0.85	12.3	1.00
7. Homogeneous RC Model (RC_1)	18	24.7	0.72	11.9	1.00
8. Heterogeneous RC Model (RC_2)	12	11.1	0.96	6.2	1.00
9. Simple Heterogeneous RC Model (RC_3)	17	16.1	0.93	9.3	1.00
10. Homogeneous Topological Model (TOP_1)	20	53.3	0.06	16.7	0.99
11. Heterogeneous Topological Model (TOP_2)	16	42.1	0.11	12.0	0.98
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	45.5	0.12	14.2	0.98

Table 11

Condition 10: Small Difference in Association, Large Difference in Group Size

Model Description	*df*	L^2	*p*	L^2_s	*p*
(a) LS1R: N_1=1,000, ϕ_1=1.5; N_2=6,000, ϕ_2=1.0					
1. Conditional Independence Model (CI)	24	286.0	0.00	92.9	0.00
2. Full Two-Way Interaction Model (FI)	12	16.0	0.84	10.2	0.99
3. Log-Linear Layer Effect Model (LL_1)	11	11.1	0.94	6.9	1.00
4. Log-Multiplicative Layer Effect Model (LL_2)	11	11.1	0.95	5.1	1.00
5. Homogeneous UA Model (UA_1)	23	29.1	0.80	16.6	0.99
6. Heterogeneous UA Model (UA_2)	22	24.8	0.87	13.6	0.99
7. Homogeneous RC Model (RC_1)	18	22.0	0.87	13.6	0.99
8. Heterogeneous RC Model (RC_2)	12	11.7	0.96	6.7	0.99
9. Simple Heterogeneous RC Model (RC_3)	17	17.1	0.95	9.4	0.99
10. Homogeneous Topological Model (TOP_1)	20	33.2	0.50	17.6	0.97
11. Heterogeneous Topological Model (TOP_2)	16	25.9	0.57	12.7	0.98
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	29.0	0.58	14.2	0.98
(b) LS2R: N_1=6,000, ϕ_1=1.5; N_2=1,000, ϕ_2=1.0					
1. Conditional Independence Model (CI)	24	315.6	0.00	92.7	0.00
2. Full Two-Way Interaction Model (FI)	12	18.8	0.65	9.8	1.00
3. Log-Linear Layer Effect Model (LL_1)	11	11.8	0.94	6.6	1.00
4. Log-Multiplicative Layer Effect Model (LL_2)	11	11.7	0.93	8.0	1.00
5. Homogeneous UA Model (UA_1)	23	31.7	0.72	16.9	1.00
6. Heterogeneous UA Model (UA_2)	22	25.9	0.87	14.2	1.00
7. Homogeneous RC Model (RC_1)	18	25.0	0.74	13.3	1.00
8. Heterogeneous RC Model (RC_2)	12	12.1	0.96	7.0	1.00
9. Simple Heterogeneous RC Model (RC_3)	17	17.8	0.93	10.8	1.00
10. Homogeneous Topological Model (TOP_1)	20	40.6	0.21	17.7	1.00
11. Heterogeneous Topological Model (TOP_2)	16	31.1	0.34	13.1	0.98
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	34.4	0.35	15.4	0.99

Table 12
Condition 11: Small Difference in Association, Small Difference in Group Size

Model Description	*df*	L^2	*p*	L_s^2	*p*
(a) SS1R: N_1=1,000, ϕ_1=1.5; N_2=2,000, ϕ_2=1.0					
1. Conditional Independence Model (CI)	24	136.3	0.00	95.2	0.00
2. Full Two-Way Interaction Model (FI)	12	15.8	0.77	12.4	0.94
3. Log-Linear Layer Effect Model (LL_1)	11	11.5	0.91	8.9	0.98
4. Log-Multiplicative Layer Effect Model (LL_2)	11	11.2	0.95	7.5	1.00
5. Homogeneous UA Model (UA_1)	23	27.2	0.85	20.8	0.98
6. Heterogeneous UA Model (UA_2)	22	23.5	0.94	17.9	0.99
7. Homogeneous RC Model (RC_1)	18	21.7	0.84	16.8	0.97
8. Heterogeneous RC Model (RC_2)	12	11.8	0.95	8.9	0.99
9. Simple Heterogeneous RC Model (RC_3)	17	17.2	0.92	12.6	1.00
10. Homogeneous Topological Model (TOP_1)	20	28.1	0.69	21.4	0.89
11. Heterogeneous Topological Model (TOP_2)	16	21.3	0.74	15.9	0.92
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	24.0	0.77	17.8	0.96
(b) SS2R: N_1=2,000, ϕ_1=1.5; N_2=1,000, ϕ_2=1.0					
1. Conditional Independence Model (CI)	24	146.8	0.00	96.7	0.00
2. Full Two-Way Interaction Model (FI)	12	17.6	0.74	12.5	0.96
3. Log-Linear Layer Effect Model (LL_1)	11	11.6	0.94	8.5	0.97
4. Log-Multiplicative Layer Effect Model (LL_2)	11	11.2	0.94	9.2	0.97
5. Homogeneous UA Model (UA_1)	23	29.8	0.75	21.4	0.99
6. Heterogeneous UA Model (UA_2)	22	24.5	0.92	17.8	1.00
7. Homogeneous RC Model (RC_1)	18	24.3	0.78	17.3	0.95
8. Heterogeneous RC Model (RC_2)	12	12.7	0.95	9.3	0.99
9. Simple Heterogeneous RC Model (RC_3)	17	18.1	0.95	13.7	0.99
10. Homogeneous Topological Model (TOP_1)	20	31.0	0.54	21.6	0.93
11. Heterogeneous Topological Model (TOP_2)	16	22.5	0.75	15.7	0.98
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	25.6	0.74	18.2	0.98

Table 13
Condition 12: Small Difference in Association, No Difference in Group Size

Model Description	*df*	L^2	*p*
(a) NS1R: N_1=1,000, ϕ_1=1.5; N_2=1,000, ϕ_2=1.0			
1. Conditional Independence Model (CI)	24	102.7	0.00
2. Full Two-Way Interaction Model (FI)	12	15.6	0.82
3. Log-Linear Layer Effect Model (LL_1)	11	11.4	0.97
4. Log-Multiplicative Layer Effect Model (LL_2)	11	11.1	0.98
5. Homogeneous UA Model (UA_1)	23	27.8	0.84
6. Heterogeneous UA Model (UA_2)	22	24.3	0.93
7. Homogeneous RC Model (RC_1)	18	22.1	0.82
8. Heterogeneous RC Model (RC_2)	12	12.2	0.95
9. Simple Heterogeneous RC Model (RC_3)	17	17.8	0.97
10. Homogeneous Topological Model (TOP_1)	20	27.2	0.75
11. Heterogeneous Topological Model (TOP_2)	16	20.6	0.82
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	23.5	0.84

Table 14
Condition 13: No Difference in Association, Huge Difference in Group Size

Model Description	*df*	L^2	*p*	L_s^2	*p*
(a) HN1R: N_1=1,000, ϕ_1=1.0; N_2=12,000, ϕ_2=1.0					
1. Conditional Independence Model (CI)	24	477.7	0.00	68.6	0.00
2. Full Two-Way Interaction Model (FI)	12	12.4	0.96	7.1	1.00
3. Log-Linear Layer Effect Model (LL_1)	11	11.1	0.96	6.3	1.00
4. Log-Multiplicative Layer Effect Model (LL_2)	11	11.1	0.96	6.0	1.00
5. Homogeneous UA Model (UA_1)	23	27.6	0.80	13.4	1.00
6. Heterogeneous UA Model (UA_2)	22	26.3	0.83	12.5	1.00
7. Homogeneous RC Model (RC_1)	18	18.2	0.96	9.6	1.00
8. Heterogeneous RC Model (RC_2)	12	11.2	0.99	6.0	1.00
9. Simple Heterogeneous RC Model (RC_3)	17	16.9	0.96	10.4	1.00
10. Homogeneous Topological Model (TOP_1)	20	35.4	0.37	13.3	1.00
11. Heterogeneous Topological Model (TOP_2)	16	31.2	0.33	10.6	1.00
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	34.0	0.35	12.3	1.00

Table 15
Condition 14: No Difference in Association, Large Difference in Group Size

Model Description	*df*	L^2	*p*	L_s^2	*p*
(a) LN1R: N_1=1,000, ϕ_1=1.0; N_2=6,000, ϕ_2=1.0					
1. Conditional Independence Model (CI)	24	263.7	0.00	70.5	0.00
2. Full Two-Way Interaction Model (FI)	12	10.9	0.97	6.8	1.00
3. Log-Linear Layer Effect Model (LL_1)	11	10.1	0.98	6.3	1.00
4. Log-Multiplicative Layer Effect Model (LL_2)	11	10.0	0.97	5.7	1.00
5. Homogeneous UA Model (UA_1)	23	24.2	0.94	13.1	1.00
6. Heterogeneous UA Model (UA_2)	22	23.5	0.90	12.6	1.00
7. Homogeneous RC Model (RC_1)	18	16.6	0.99	9.8	1.00
8. Heterogeneous RC Model (RC_2)	12	10.6	0.99	5.8	1.00
9. Simple Heterogeneous RC Model (RC_3)	17	15.8	0.99	9.0	1.00
10. Homogeneous Topological Model (TOP_1)	20	27.7	0.60	13.1	1.00
11. Heterogeneous Topological Model (TOP_2)	16	24.1	0.61	10.8	1.00
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	26.9	0.61	12.5	1.00

Table 16
Condition 15: No Difference in Association, Small Difference in Group Size

Model Description	*df*	L^2	*p*	L^2_s	*p*
(a) SN1R: N_1=1,000, ϕ_1=1.0; N_2=2,000, ϕ_2=1.0					
1. Conditional Independence Model (CI)	24	116.7	0.00	73.2	0.00
2. Full Two-Way Interaction Model (FI)	12	11.8	0.94	9.2	1.00
3. Log-Linear Layer Effect Model (LL_1)	11	10.7	0.94	8.3	1.00
4. Log-Multiplicative Layer Effect Model (LL_2)	11	10.7	0.95	8.2	0.99
5. Homogeneous UA Model (UA_1)	23	24.0	0.94	17.7	1.00
6. Heterogeneous UA Model (UA_2)	22	22.9	0.93	16.8	1.00
7. Homogeneous RC Model (RC_1)	18	18.0	0.94	13.7	1.00
8. Heterogeneous RC Model (RC_2)	12	11.6	0.94	8.5	1.00
9. Simple Heterogeneous RC Model (RC_3)	17	16.9	0.95	12.9	1.00
10. Homogeneous Topological Model (TOP_1)	20	23.2	0.83	16.8	1.00
11. Heterogeneous Topological Model (TOP_2)	16	18.9	0.89	13.4	0.99
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	22.1	0.85	15.9	0.99

Table 17
Condition 16: No Difference in Association, No Difference in Group Size

Model Description	*df*	L^2	*p*
(a) NN1R: N_1=1,000, ϕ_1=1.0; N_2=1,000, ϕ_2=1.0			
1. Conditional Independence Model (CI)	24	77.4	0.00
2. Full Two-Way Interaction Model (FI)	12	12.4	0.93
3. Log-Linear Layer Effect Model (LL_1)	11	11.3	0.93
4. Log-Multiplicative Layer Effect Model (LL_2)	11	11.1	0.91
5. Homogeneous UA Model (UA_1)	23	23.6	0.95
6. Heterogeneous UA Model (UA_2)	22	22.5	0.95
7. Homogeneous RC Model (RC_1)	18	18.0	0.93
8. Heterogeneous RC Model (RC_2)	12	11.3	0.96
9. Simple Heterogeneous RC Model (RC_3)	17	16.7	0.93
10. Homogeneous Topological Model (TOP_1)	20	22.2	0.91
11. Heterogeneous Topological Model (TOP_2)	16	18.1	0.89
12. Topological Model with Log-Multiplicative Layer Effect (TOP_3)	19	21.0	0.92

Table 18
Summary of the Usefulness of Chi-Square Difference Test to Detect Group Differences

	Δ_1	Δ_{1s}	Δ_2	Δ_{2s}	Δ_3	Δ_{3s}	Δ_4	Δ_{4s}	Δ_5	Δ_{5s}	Δ_6	Δ_{6s}	Δ_7	Δ_{7s}
HH1R	100	100	100	100	100	100	100	100	100	100	100	100	100	100
HH2R	100	100	100	100	100	100	100	100	100	100	100	100	100	100
LH1R	100	100	100	100	100	100	100	100	100	100	100	100	100	100
LH2R	100	100	100	100	100	100	100	100	100	100	100	100	100	100
SH1R	100	100	100	100	100	100	100	100	100	100	100	100	100	100
SH2R	100	100	100	100	100	100	100	100	100	100	100	100	100	100
NH1R	100	---	100	---	100	---	100	---	100	---	100	---	100	---
HL1R	97	92	96	96	94	89	84	61	96	95	84	67	95	92
HL2R	100	98	100	86	100	97	97	52	100	92	98	65	100	96
LL1R	99	95	97	100	100	95	88	62	99	99	88	71	95	94
LL2R	100	87	100	87	100	96	97	54	100	95	99	67	100	96
SL1R	95	93	94	96	94	93	79	68	96	96	86	80	94	93
SL2R	99	96	98	96	98	95	89	65	99	95	91	77	99	95
NL1R	89	---	86	---	87	---	65	---	89	---	66	---	83	---
HS1R	55	39	54	63	51	35	30	8	54	51	30	5	49	37
HS2R	76	35	77	14	68	25	44	3	75	23	54	5	74	16
LS1R	47	30	56	62	42	32	26	9	51	47	29	8	44	38
LS2R	74	30	72	13	66	24	43	4	74	21	51	5	68	20
SS1R	50	40	54	57	39	26	26	13	49	47	25	12	41	37
SS2R	60	45	63	35	50	44	40	14	63	40	41	16	52	34
NS1R	39	---	48	---	32	---	24	---	43	---	15	---	35	---
HN1R	10	4	11	7	10	7	11	0	9	4	7	0	9	4
LN1R	2	1	4	4	3	0	8	0	2	1	4	0	2	1
SN1R	7	4	8	6	6	3	11	6	6	5	8	3	7	4
NN1R	7	---	10	---	9	---	7	---	5	---	3	---	9	---

Note: Δ_1=L^2 difference between FI & LL_1; Δ_2=L^2 difference between FI & LL_2; Δ_3=L^2 difference between UA_1 & UA_2; Δ_4=L^2difference between RC_1 & RC_2; Δ_5=L^2 difference between RC_1 & RC_3; Δ_6=L^2 difference between TOP_1 & TOP_2; Δ_7=L^2 difference between TOP_1 & TOP_3; and Δ_{1s}, Δ_{2s}, Δ_{3s}, Δ_{4s}, Δ_{5s}, Δ_{6s}, and Δ_{7s} refer to the same L^2 difference but with standardization. See text for details