Published in final edited form as:

Struct Equ Modeling. 2017 ; 24(4): 626–635. doi:10.1080/10705511.2016.1276837.

Structural Models for Binary Repeated Measures: Linking Modern Longitudinal Structural Equation Models to Conventional Categorical Data Analysis for Matched Pairs

Jason T. Newsom

Abstract

The current widespread availability of software packages with estimation features for testing structural equation models with binary indicators makes it possible to investigate many hypotheses about differences in proportions over time that are typically only tested with conventional categorical data analyses for matched pairs or repeated measures, such as McNemar's chi-squared. The connection between these conventional tests and simple longitudinal structural equation models is described. The equivalence of several conventional analyses and structural equation models reveals some foundational concepts underlying common longitudinal modeling strategies and brings to light a number of possible modeling extensions that will allow investigators to pursue more complex research questions involving multiple repeated proportion contrasts, mixed between-subjects $\times$ within-subjects interactions, and comparisons of estimated membership proportions using latent class factors with multiple indicators. Several models are illustrated, and the implications for using structural equation models for comparing binary repeated measures or matched pairs are discussed.

Sophisticated structural equation models for investigating change in longitudinal studies, such as latent growth curve models and latent difference score models, are becoming increasingly common, yet there is limited consideration of how such models relate to conventional analyses for repeated measures. Greater exploration of the fundamental statistical concepts underlying simple structural equation models for repeated measures opens of a variety of new avenues for the analysis of change, however. Several simple structural equation models that examine mean change in continuous variables are equivalent to repeated measures ANOVA under specific conditions (Newsom, 2002; 2015; Rovine & Molenaar, 2003; Voelkle, 2007) and are core components of models widely used to investigate individual change. An examination of how the same models can be used with binary or other discrete variable types reveals a similar connection to conventional categorical data analyses as well as many potentially valuable modeling extensions and statistical advantages. In fact, because the same specifications can be used for continuous

Portland State University, Psychology, P.O. Box 751, Department of Psychology, Portland, OH 97207.

The data for the examples in this paper were obtained with funding from the National Institutes of Health R01 AG14130. I thank Alan Agresti for comments on a portion of an earlier draft of this paper and participants of the Portland State University Stats-Lunch Brown Bag Series and the 2015 Modern Modeling Methods Conference for helpful suggestions on this work. Correspondence should be addressed to Jason T. Newsom, Department of Psychology, Portland State University, P.O. Box 751, Portland, OR 97207-0751.

and binary variables, it is clear that structural modeling provides a powerful common framework for analyzing repeated measures regardless of the type of observed variable.

With the generalization of structural equation models that include binary and ordinal dependent variables under well-performing estimation approaches (Muthén, 2001), nearly all structural model specifications for longitudinal analysis of continuous variables can be applied to categorical dependent variables. Use of the common latent growth curve model specification with binary variables is one example (e.g., Masyn, Petras, & Liu, 2014; Muthén, 1996), but there has been relatively infrequent discussion of the application of other longitudinal modeling strategies to categorical data. One such application and some of its potential extensions will be considered—structural equation models for comparing proportions over two time points. An exploration of this topic not only helps elucidate the underpinnings of more complex structural models with discrete variables, but it illustrates how a general structural equation modeling framework can transcend conventional categorical analysis by incorporating ordinal or other variable types, robust estimation methods, advanced missing data handling, multigroup analysis, and latent class modeling with multiple indicators for the analysis of any combination of between-subjects and within-subjects factor effects.

Continuous Variable Case: Repeated Measures ANOVA

To understand models for comparing repeated proportions in a generalized modeling framework, it is useful to first briefly review structural equation modeling approaches to comparing means for repeated measurements (a more elaborate discussion can be found in Newsom, 2015). As illustrated in Figure 1, there are several modeling specifications that can be shown to be equivalent to a repeated measures ANOVA, including a model with a single gain score computed from two measurements, $y_2 - y_1$ (Rovine & Molenaar, 2003), a model with two correlated measured variables, a simple latent growth curve model (Newsom, 2002; Voelkle, 2007), or a simple latent difference score model (McArdle, 2001). When specified with a mean structure, all of the models allow for a test of the repeated measures mean comparison equivalent or asymptotically equivalent to repeated measures ANOVA assuming complete data (Newsom, 2015). For each of the models, except the model Figure 1b, a test of the hypothesis that the means at Time 1 and Time 2 differ significantly can be carried out by a Wald test or a likelihood ratio test of the difference between two factor means (given as α_2 in Figure 1a, 1c, or 1d). The likelihood ratio test requires a second model which constrains the mean to be equal to 0. The Wald and likelihood ratio tests may differ depending on various conditions (Satorra, 1989), but they will be asymptotically equivalent under ideal circumstances. The hypothesis test of mean differences for the model illustrated in Figure 1b can only be carried out through the likelihood ratio test comparing models within and without the means (α_1 and α_2) constrained to be equal. It can be shown, for example, that α_2 for the simple growth curve model depicted in Figure 1c is equal to the same mean difference obtained through repeated measures ANOVA (Newsom, 2015).

$$E(y_{ti}) = \alpha_1 + \alpha_2 \quad (1)$$

In the above equation, α_1 represents the intercept or Time 1 mean, whereas α_2 represents the difference between the Time 1 and Time 2 means. An individual index subscript is not given in the equation in this instance, but either factor could be specified as a random effect by estimating the variance of the factor.¹ Estimating the difference factor variance makes the model equivalent to a random effects repeated-measures ANOVA (Newsom, 2002).

Models shown in Figure 1b through 1d also easily can be extended to the multiple indicator case (see Newsom, 2015; Voelkle, 2008). There is no statistical reason to prefer one of these model specifications over the other, but for convenience the subsequent discussion of binary and ordinal variables will be confined to the simple growth curve model depicted in Figure 1c, hereafter called the “contrast coding model.”

Comparing Proportions at Two Time Points

Conventional tests.

The simplest case is the investigation of change in a single measured binary variable over two time points, usually referred to more generally as a matched pairs comparison. To make matters more concrete, examples will be discussed in terms of repeated-measures, such as examining the change in proportions in a survey with a binary outcome (e.g., change in endorsement of “yes” over time in a political survey that is re-administered to the same respondents), but these models apply equally to other dependent comparisons involving dyads, matched cases in experimental and quasi-experimental designs, or multiple observers. The most common conventional test for the matched pairs comparison of a binary response is the McNemar’s chi-squared (McNemar, 1947). The McNemar test is equivalent to the sign test, Cochran’s Q extends the test to more than two time points, and Liddell’s test is an exact test version (Agresti, 2013). Like the Pearson chi-squared test, the McNemar is a form of score test because the standard error is based on the proportions expected under the null hypothesis (Klingenberg & Agresti, 2006).

The McNemar statistic can be equivalently stated in terms of the difference between marginal proportions (“yes” vs. “yes” or “no” vs. “no”) for a 2×2 repeated measures table or in terms of the discordant cells (“yes”-“no” vs. “no”-“yes”). If $p_{1.}$ and $p_{.1}$ refer to the row and column proportions, respectively, and p_{12} and p_{21} refer to the discordant cell proportions, where each $p_{jk} = n_{jk}/n$, for row j and column k .

$$p_{1.} - p_{.1} = (p_{11} + p_{12}) - (p_{11} + p_{21}) = p_{12} - p_{21} \quad (2)$$

Thus, the test of the equality of marginal proportions (marginal homogeneity) is also a test of discordant cell proportions in the binary two-variable case

¹With two time points, it is not possible to estimate the variances of both the intercept and the slope as well as unique estimates of measurement residual variance at each time point (Newsom, 2002; Voelkle, 2007).

$$McNemar's \chi^2 = \frac{(p_{1.} - p_{.1})^2}{\frac{1}{n}(p_{1.} + p_{.1} - 2p_{11})} = \frac{(p_{12} - p_{21})^2}{p_{12} + p_{21}} \quad (3)$$

The McNemar chi-squared is a special case of the Stuart-Maxwell statistic W for a larger $J \times K$ matched-pair contingency table (Stuart 1955), where the number of rows J and columns K are equal.

$$W = n\mathbf{d}'\mathbf{V}^{-1}\mathbf{d} \quad (4)$$

This matrix formulation includes the scalar, n , for sample size, $\mathbf{d}$, a vector of discordant cell proportions, and $\mathbf{V}$, a lower triangular matrix, with diagonal elements, V_{jj} , equal to $n_{j.} + n_{.j} - 2n_{jj}$ and below-diagonal elements, V_{jk} , equal to $-(n_{jk} + n_{kj})$, where subscript jj denotes that $j = k$.

When the McNemar test is stated in terms of z -test rather than chi-squared test, the square root of the denominator just to the right of the first equal sign in Equation (3), $\sqrt{\frac{1}{n}(p_{1.} + p_{.1} - 2p_{11})}$, is the standard error. The standard error also can be obtained from the raw data matrix $\mathbf{Y}$ consisting of observed 0 and 1 responses at the two time points, y_1 and y_2 . For binary values, the sum of squares cross-product matrix (SSCP) contains marginal proportions and the cell proportion p_{11} (e.g., proportion responding “yes” at Time 1 and “yes” at Time 2).

$$SE = \left(\frac{1}{n^2}\right)\mathbf{1}'[(\mathbf{Y}'\mathbf{Y})^{-1}|\mathbf{Y}'\mathbf{Y}|\mathbf{1}]$$

The sum of the elements in the square bracketed term is the quantity $p_{1.} + p_{.1} - 2p_{11}$ for the binary case, which is, incidentally, the same as the adjoint matrix, $adj(\mathbf{Y})$.

A conditional logistic model can also be used for binary matched pairs (Cox, 1958; Rasch, 1961). The model is tested by regressing time-specific scores on the outcome, y_{it} , on the time metric variable (e.g., $x_t = 0, 1$) using a long-format data structure and standard logistic regression.

$$\text{logit}[P(y_{it} = 1)] = \alpha_i + \beta x_t \quad (5)$$

Where the intercept α_i varies across cases, the slope is conditioned on subject; and the result is a subject-specific effect. The McNemar chi-squared test, on the other hand, tests a population-averaged effect. The two types of tests do not always lead to the same

conclusions as a result. The coefficient from the conditional logistic regression is obtainable from the log of the discordant cell proportions (Agresti, 2013).

$$\hat{\beta} = \ln\left(\frac{p_{21}}{p_{12}}\right) \quad (6)$$

The coefficient, $\hat{\beta}$, no longer captures a comparison of the marginal proportions, because $\ln(p_{21}/p_{12}) \neq \ln(p_{1.}/p_{.1})$. The natural log of the ratio of proportions can be restated as the difference in the logs of each proportion, where $\ln(p_{21}/p_{12}) = \ln(p_{21}) - \ln(p_{12})$, using the quotient rule. With the logit link, the difference in the logs of the discordant cell proportions is not the same as the difference in the logs of the marginal proportions, however, $\ln(p_{21}) - \ln(p_{12}) \neq \ln(p_{1.}) - \ln(p_{.1})$.² For ordinary least squares regression (OLS), the link is identity and the average of the difference scores is equal to the difference in the averages, so there is no distinction between subject-specific and population-averaged effect. For example, the test of whether two dependent means are equivalent is the same as the test of whether the average difference score is equal to 0 (e.g., using a dependent *t*-test or repeated measures ANOVA).

Structural Equation Modeling Tests for Change in a Binary Variable

Simple structural equation models, as represented in Figure 1b, 1c, or 1d, can be specified with a binary dependent variable to provide tests of repeated measures differences. Though the same general concepts apply to each of these specifications, for the sake of brevity, consider just the contrast coding model based on the latent growth curve specification that is depicted in Figure 2. With the specification shown, the model is just identified.

Marginal maximum likelihood (ML; Bock & Aitkin, 1981) and diagonal weighted least squares with robust standard error estimates (DWLS; Muthén, du Toit, & Spisic, 1997) are two common approaches that could be used for estimation of this model. Other estimation approaches are possible but will not be addressed in the present discussion. Inclusion of the mean structure in estimating SEMs with binary variables will produce a factor mean estimate for α_2 that corresponds to the appropriate cumulative distribution function (cdf) link to the model-based proportion estimates. Most commonly, the ML estimate implies a logit link function, whereas the DWLS with theta parameterization implies a probit link function. An equation for the contrast coding model is parallel to the conditional logistic analysis model in Equation (5) if the logit link function is used:

$$\text{logit}[P(y_{it} = 1)] = \alpha_{1i} + \alpha_2 \lambda_{t2}$$

²Because $p_{21} - p_{12} = p_{1.} - p_{.1} = \sum (y_{2i} - y_{1i})/n$, the results from taking the log of each of these quantities are also equal, so that $\ln(p_{21} - p_{12}) = \ln(p_{1.} - p_{.1}) = \ln\left[\sum (y_{2i} - y_{1i})/n\right]$.

To be parallel to the conditional logistic model, the structural model would freely estimate the variance of the intercept factor, $Var(\eta_1) = \psi_{11}$, but not the slope factor, $Var(\eta_2) = \psi_{22} = 0$. In multilevel terms, the intercept value, α_{1j} , is allowed to vary across cases but the slope value, α_2 , is non-varying.

Wald Test

Because the slope for conditional logistic model is a function of the discordant cell proportions, as shown in Equation (6) for the conditional logistic model, it follows that the slope mean for the contrast coding model is also a function of the discordant cell proportions. And, for ML with logit-link function, the factor mean estimate similarly would be a function of the natural log of the discordant cell proportions:

$$\alpha_2 = \ln\left(\frac{p_{21}}{p_{12}}\right)$$

The test of $\alpha_2 = 0$ using the Wald z-test with ML estimation then is a test of the subject-specific effect of the change in the binary response that is equal to the Wald z-test of the time coefficient, $\hat{\beta}/SE_{\hat{\beta}}$, for the conditional logistic regression, where $\hat{\beta}$ is equal to Equation (6) and the standard error is equal to $\sqrt{(1/np_{21}) + (1/np_{12})}$.

Nested Model Test

A nested model test also can be used to investigate change in the binary response. Because the contrast coding model (as well as the other equivalent models) is just identified, the fit of a model constraining the slope factor mean, α_2 , to be equal to 0, is a single degree of freedom test of the two nested models. The interpretation of this test differs from the interpretation of the Wald test, however. Both a Pearson chi-squared and likelihood ratio chi-squared have been used for overall model fit with binary factor analysis with ML (Muthén, 1993). The fit indices for binary factor models are derived from a function of the discrepancy between expected and observed residuals using the multi-way 2^q margin data table for q binary variables (Christofferson, 1975; Muthén & Christofferson, 1981; Muthén, 1978) minimizing

$$n(\mathbf{p}_o - \mathbf{p}_e)' \mathbf{S}^{-1} (\mathbf{p}_o - \mathbf{p}_e)$$

The Pearson chi-squared used for model fit can be obtained and follows its usual computation

$$\chi^2 = \sum_1^{2q} \frac{(f_O - f_E)^2}{f_E}$$

q is the dimension of the contingency table (i.e., the number of variables), f_O is the observed frequency and f_E is the expected frequency. Constraint of α_2 to be equal to 0 implies that the

expected marginal proportions are equal and, therefore, according the Equation (2), also that the discordant cell proportions are equal. It is not surprising then that the Pearson chi-squared for the model fit of the constrained model produces a value equal to McNemar's chi-squared test for equality of proportions in the repeated measures case.³

The likelihood ratio goodness-of-fit index also can be computed. The likelihood ratio approach uses a natural log function of the observed and expected frequencies.

$$G^2 = 2 \sum_{Y=1}^{2^k} f_O \ln \frac{f_O}{f_E}$$

Both fit indices imply a degrees of freedom equal to $2^q - q(q+1) - 1$ in which 2^q is the number of possible pairwise elements. The likelihood ratio for the model fit of the constrained model produces a value equal to the likelihood ratio for the fit of the conventional conditional logistic model, a function of the ratio of the reduced to the full model. For conditional logistic regression for the repeated measures case, the reduced model contains only the intercept and the full model contains the intercept and the time predictor.

For either fit index, the expected frequencies implied by the model are simply a function of the specified model parameters in the usual manner. In general, the equation for the model implied means (e.g., Mulaik, 2009) is

$$\alpha = \bar{Y}(\Lambda' \Lambda)^{-1} \Lambda'$$

Λ is the loading matrix as given above and shown in Figure 2 and all measurement intercepts, ν , are 0. Lack of fit in this circumstance is a function of the differences in mean structures implied by the two models. The constrained model imposes α_2 to be equal to 0 in the mean vector, but all other elements of the implied variance-covariance matrix are the same in the two models. Estimates from the unconstrained model, which is just identified for the model in Figure 2, include an intercept mean, α_1 , equal to the log of the expected marginal proportion at Time 1 and a contrast factor mean, α_2 , equal to the log of the average difference of the binary scores.

³Although there is a striking similarity in Equations (4) for Stuart's (1955) generalization of McNemar's chi-squared and Equation (7) for the Christofferson's (1975) fit function for binary factor models, the elements of each are different. Both quadratic forms take into account the sample size and include a cross product of proportions and inverse of a covariance matrix. Stuart's equation involves a vector of proportions of the discordant cells, $\mathbf{d}$, as well as the inverse of the null variance-covariance matrix, which is a lower triangular matrix of expected values for the marginal homogeneity hypothesis. Christofferson's equation involves a vector of expected proportions minus observed proportions, $\mathbf{p}_e - \mathbf{p}_o$, that are functions of the loading and factor variance-covariances implied by the model as well as the inverse of the variance-covariance matrix of proportions. As described above, the variance-covariance matrix of proportions will contain information about the expected marginal and cell proportions. The similarity is far from coincidental, therefore, because, depending on the specifications for the structural model, the fit obtained by minimizing Christofferson's equation can be a test of the discordant proportions involved in the homogeneity test.

Extensions

There are several possible extensions of the simple two-time point contrast coding model. One extension is to add additional time points, with $T - 1$ contrast factors and a single intercept. The loading matrix can be configured to test a variety of hypotheses that contrast proportions at different time points (e.g., backward difference coding, Helmert coding, deviation coding). For example, a model, such as the one illustrated in Figure 3 could be used to test whether there is a difference in the proportion of cases exceeding a clinical cutoff of depression six months and a year as compared with the baseline measurement. But the coding scheme can be modified to make comparisons between the six-month follow-up and the baseline and the one-year follow-up and baseline. When the comparisons are between adjacent time points, the model is equivalent to the latent change score model (McArdle, 2001).

Additional variables can be included to investigate whether latent or measured variables are predictors of the repeated measures differences or are predicted by the repeated measures differences. For a model that includes one or more predictors, following the general multiple indicator multiple cause (MIMIC) strategy, the path between the predictor and the contrast factor is the same as a cross-level interaction in growth curve analysis. In the presence of a predictor, the estimate of the contrast factor mean (i.e., intercept) is a conditional mean that can provide information about the adjusted proportion change using the logistic transformation. A special case of this model when the predictor is binary can provide a test of a mixed between-subjects $\times$ within-subjects interaction for a binary response variable difficult to otherwise obtain with conventional software. Alternatively, moderator hypotheses can be investigated using multigroup models, where constraints on the intercept or contrast factor provide between-group tests of baseline proportions or proportional change.

The contrast coding model also can be extended by using categorical latent variables rather than observed binary variables as indicators. Latent class models (Lazarsfeld & Henry, 1968; Goodman, 1974) are used to assign probability of membership to cases in a data set according to a set of binary or continuous indicators. For continuous indicators, the model is known as a latent profile analysis (Lazarsfeld & Henry, 1968). Either model can be considered as a categorical latent variable, and, when combined with path analytic components or continuous latent variables fits these finite mixture models into a more generalized conception of structural equation modeling (Lubke & Muthén, 2005; Muthén, 2001). A contrast coding model with latent class indicators is a second-order mixture factor model. Such a model is closely related to latent transition models (Collins & Lanza, 2010), although the latent transition model examines the probability that an individual will transition from one class to another over time rather than an estimate of the average subject-specific change over time or difference in proportions.

Examples

To illustrate the test of change in a binary variable for two time points and the equivalence of the contrast coding model to conventional tests, depression scores from the Late Life Study of Social Exchanges (LLSSE; Sorkin & Rook, 2004) were used. The depression scores are

binary variables derived from a brief 9-item version (Santor & Coyne, 1997) of the Center for Epidemiologic Studies-Depression scale (Radloff, 1977). The LLSSE study is a longitudinal study of social relationships among adults 65–90 years of age. The initial analysis focused only on the first two assessments taken six months apart, and, for simplifying comparisons between various models, any cases with missing data were eliminated, resulting in a total $N = 719$. For the first example, binary variables were computed for each time point by grouping those with a score of four or higher on the CES-D, a cutoff shown by Santor and Coyne to have optimal sensitivity and specificity in predicting diagnosis of clinical depression using psychiatric interview. Dichotomizing otherwise continuous variables has potential risks of misclassification (MaCallum, Zhang, Preacher, & Rucker, 2002), but a comparison of clinical cutoffs may have therapeutic value in some instances and the issue of classification error potentially can be at least partially addressed in the subsequent latent class analysis example.

Contrast Coding Model for Two Binary Repeated Measures

The proportions of those above and below the depression threshold were compared at the two time points. Three analysis methods were used to illustrate. The structural equation modeling approach specified the contrast coding model just as illustrated in Figure 2, using ML and logistic parameterization for binary variables in Mplus 7.4 (Muthén & Muthén, 1998–2015). Two versions of the model were tested, however, one in which the factor mean for the contrast factor, α_2 , was freely estimated and one in which it was constrained to be equal to 0. The model fit information then represents a test of the whether there was significant change in the binary variable over time. Conventional analyses were used for comparison using SAS Version 9.4. McNemar's chi-squared was obtained using PROC FREQ and /AGREE option. The conditional logistic model was estimated by first converting the data to long format with one record per observation and then using PROC LOGISTIC and the STRATA statement for individual identification numbers.

Table 1 gives the percentage of cases scoring in the clinical depression range at Time 1 and Time 2. At Time 1, there were 383 cases (53.27%) and at Time 2, there were 505 (70.24%) in the depressed category. The unconstrained model, which was just identified, had an intercept factor mean, α_1 , equal to .223, $p = .082$. The estimate for the contrast factor mean was significant, α_2 , equal to 1.222, $SE = .159$, $p < .001$ according the Wald ratio test, suggesting a significant increase in the probability that an individual scored in the clinical depression range. The factor means can be used to estimate the proportion of variance at baseline or the proportional difference using the logistic cdf. For the intercept factor mean estimate α_1 , the logistic cdf computation estimates that 53.27% of the sample scored in the depressed range,

$$\begin{aligned} & \exp(\alpha_1) / [1 + \exp(\alpha_1)] \\ &= \exp(.223) / [1 + \exp(.223)] \\ &= .5327 \end{aligned}$$

and, using the factor mean estimate for α_2 , the contrast factor, the logistic cdf computation estimates that an increase of approximately 17.7% in the probability of being classified as depressed

$$\begin{aligned} & \left[\exp(\alpha_1 + \alpha_2) / [1 + \exp(\alpha_1 + \alpha_2)] \right] - \left[\exp(\alpha_1) / [1 + \exp(\alpha_1)] \right] \\ &= [\exp(.223 + 1.222) / [1 + \exp(.223 + 1.222)]] - [\exp(.223) / [1 + \exp(.223)]] \\ &= .177 \end{aligned}$$

Note that this value differs somewhat from the difference in marginal proportions, because logistic cdf estimate of the contrast factor mean represents the subject-specific (or conditional) effect rather than the population-averaged (or marginal) effect.

The constrained version of the contrast coding model, with α_2 set equal to 0, indicated a significant difference in the proportion of cases with scores in the clinical depression range, with Pearson $\chi^2(1) = 66.448$, $p < .001$. The results from model fit and test of the Wald test of the significance of the contrast factor mean are not equal here, $(1.222/.159)^2 = (7.666)^2 = 58.768$, because they represent population-averaged and subject-specific tests, respectively.

The conventional tests closely matched the structural model results. The McNemar's chi-squared, the test of the equivalence of the marginal (or discordant cell) proportions, was significant, McNemar's $\chi^2(1) = 66.446$, $p < .001$, and nearly identical to the Pearson chi-squared for the fit of the constrained contrast coding model. The coefficient estimate and Wald test for time obtained with the conditional logistic analysis closely matched the values from the contrast coding model, $b = 1.222$, $SE = .159$, Wald chi-squared = 58.767.

Contrast Coding Model for Three Binary Repeated Measures

To illustrate one extension of the contrast coding model for binary variables, change in the brief CES-D cutoff scores over three waves was examined, each separated by six month intervals. For convenience and simpler comparison with conventional analyses, listwise deletion was used, resulting in a smaller sample size ($N = 651$). The model is depicted in Figure 3. To extend the model to a comparison of proportions at three time points, three factors are needed—an intercept factor and two contrast factors. Several possible contrasts could be specified, but, for this example, second time point was compared to baseline and the third time point was compared to baseline. As in the prior model, the intercept factor had loadings set equal to 1 for all three time points. The two contrast factors were specified with factor loadings accordingly to the two desired contrasts, with loadings equal to 0 1 0, and 0 0 1 at the three time points for the two contrast factors respectively. Other loadings could be specified, however, to reflect different contrasts (e.g., Time 1 vs. Time 2 and Time 2 vs. Time 3 would require loadings equal to 1 0 0 and 0 0 1). The model allows for significance tests of each individual contrast, reflecting the subject-specific difference between Time 1 and Time 2 and Time 1 and Time 3 on the binary measure. It is also possible to conducted an omnibus test by constraining the two contrast factor means to 0. As with the contrast model for two time points, although three measures provide additional degrees of freedom, there are not sufficient degrees of freedom to allow all for free estimation of all factor variances and covariances, thus only the intercept factor variance was estimated. These constraints are

not unreasonable, because they are directly analogous to a multilevel model with a random intercept but fixed regression slopes (see Newsom, 2002).

The observed proportions of those above the cutoff on the depression scale at the three time points were .538, .704, and .565, respectively, suggesting that the proportion of the sample that was depressed increased in the first six months but decreased again in the following six months. The contrast coding model enables a test of whether the second and third time points differ significantly from the first, however, which is a question that is not answered with the coding scheme used for growth curve analysis. The model allowing the intercept and both contrast factors to be freely estimated had a nonsignificant fit overall, Pearson $\chi^2(3) = 3.361$, $p < ns$. The estimate of the intercept factor, $\alpha_1 = .237$, $SE = .124$, $p = .056$, suggested that 56% of the sample was depressed at baseline, $\exp(\alpha_1)/[1 + \exp(\alpha_1)] = \exp(.241)/[1 + \exp(.241)] = .560$, which is a reasonable model-based approximation of the observed proportion. The estimates of the two factor means indicated that Time 1 and Time 2 significantly differed, but that there was no significant difference between Time 1 and Time 3, $\alpha_2 = 1.110$, $SE = .151$, $p < .001$, $\alpha_3 = .176$, $SE = .140$, $p = .209$.

An omnibus test of the equivalence of the proportions across all three time points can be obtained via a nested test in which the fit of a model constraining the two contrast factor means to be zero is compared to the model in which they are freely estimated. The constrained model fit was substantially poorer, $\chi^2(5) = 64.431$, $p < .001$, and the difference from the original model with freely estimated factor means was significant, $\chi^2(2) = 64.431 - 3.361 = 61.070$, $p < .001$. The fully constrained model result was near to the conventional Cochran's Q test of the difference in proportions among the three waves, $\chi^2(2) = 64.179$, $p < .001$.

Contrast Coding Model of Latent Classes at Two Time points

To illustrate how the contrast coding model can be extended to examine differences with latent class factors, the LLSSE data set was used to estimate the change in the probability of depression membership in two latent classes for the depression measure across two time points. The purpose was to compare change in the proportion of depressed individuals as estimated by empirical determination of more serious depression symptomatology using probability of latent class membership. The analysis cannot be directly compared to the contrast coding models illustrated above, because the prior models were based on binary classification according to pre-established clinical threshold cutoffs. Rather, in the present example, the latent class model involves a probability of classification into two or more categories of depression symptomatology empirically derived from the set of nine scale items used as indicators. Such an analysis might be useful in many circumstances in which groupings of cases are to be meaningful theoretically or have clinical utility.

Initially, separate latent models were estimated for Time 1 and Time 2 depression measures, with each model a latent profile analysis using nine continuous indicators of a latent class factor. Both models were tested against single class and fit significantly better according to likelihood ratio tests (e.g., Lo, Mendell, Rubin, 2007) and had a smaller sample-size

adjusted Bayesian Information Criteria (aBIC). These preliminary results are not presented here but are available from the author. Based on the measurement intercepts of the indicators, the two factors appear to be characterizable as clusters of depressed and non-depressed individuals.

A contrast coding model for latent classes at two time points was then estimated using maximum likelihood with integration algorithm and 50 start values,⁴ specified following the model illustrated in Figure 4. In the model, two latent class factors, denoted η_1^C and η_2^C , were used to estimate the probability of depression class membership at each time point, were indicators of two second-order factors, η_1 and η_2 , representing the intercept and the Time 2 vs. Time 1 contrast. As in the basic contrast coding model, the variance of the intercept factor was allowed to vary and the variance of the contrast factor in order to identify the second-order factors. Because the first-order latent class factors are binary, the second-order loadings are logistic regressions. For logistic parameterization, the fixed loadings must be specified in a natural log metric using large negative values (e.g., -15) to correspond to a traditional loading of 0 and large positive values (e.g., 15) to correspond to a traditional loading of 1.⁵ Thus, both loadings for the second-order intercept factor were set to 15 and the Time 1 and Time 2 loadings for second-order contrast factor were set to -15 and 15 respectively. In latent profile models, each latent class is defined by measurement intercept estimates that correspond to membership probabilities rather than traditional factor loadings (indicated in Figure 4 by dotted line arrows). The intercepts for this model were estimated while imposing longitudinal equality constraints (e.g., $\nu_{1c} = \nu_{10c}$, $\nu_{2c} = \nu_{11c}$, ..., $\nu_{9c} = \nu_{18c}$). Measurement residual values were allowed to vary across time, however.

No Pearson χ^2 or likelihood ratio test is available for mixture models. The sample adjusted Bayesian Information Criterion (BIC) was 29137.536, which, without comparison, is not particularly informative about the validity of the model. Parameter estimates and their tests can provide useful information about the data, however. The estimate of the mean of the first factor, α_1 , was equal to 1.174, which, using the logistic transformation, provides information about the probability of membership in the depressed class at the first time point,

$$\hat{\pi}_{c1} = \frac{e^{\alpha_1}}{1 + e^{\alpha_1}} = \frac{e^{1.174}}{1 + e^{1.174}} = .236$$

The variance of the intercept factor was also significant, .023, $z = 3.892$, $p < .001$, indicating that the probability of membership in the depressed class at Time 1 varied across individuals. The factor mean estimate for the contrast factor was $-.049$ and was significant, $z = -6.577$, $p < .001$. The logistic transformation of this value added to baseline factor mean, α_1 , estimates the probability of membership at Time 2,

⁴Retesting mixture models with different start values is recommended to avoid results that are based on a local minimum. After testing the present model using many start values and repeating estimation, there was no variation in results, suggesting that a global minimum was reached.

⁵Direct probability parameterization is also possible in Mplus.

$$\hat{\pi}_{c1} = \frac{e^{\alpha_1 + \alpha_2}}{1 + e^{\alpha_1 + \alpha_2}} = \frac{e^{1.174 - .049}}{1 + e^{1.174 - .049}} = .245,$$

The estimate of the proportional difference between Time 2 and Time 1 can be derived by subtracting these two numbers.

$$\hat{\pi}_{c1} = \frac{e^{\alpha_1 + \alpha_2}}{1 + e^{\alpha_1 + \alpha_2}} - \frac{e^{\alpha_1}}{1 + e^{\alpha_1}} = .009$$

suggesting a small but significant increase in the proportion estimated to be in the depressed class. Overall, the model using a latent classification approach suggested that approximately a quarter of the respondents were depressed and that the proportion who were depressed increased slightly, by approximately 1%, over a six-month period.

Discussion

Structural equation models for comparing repeated proportions may superficially appear to provide a more complicated approach to comparing proportions in repeated measures designs that can be more easily accomplished with more common approaches such as the McNemar test. In the simplest circumstances, this is true. The more general structural equation modeling framework, however, provides a variety of potential advantages, including analysis of any combination of between-subjects and within-subjects factor effects using a multigroup or MIMIC strategy, robust estimation methods, advanced missing data handling, the ability to incorporate ordinal or other variable types, and latent class modeling with multiple indicators. Although the focus has been on the repeated measures case, the models described here are equally applicable to other matched pairs contexts, such as dyadic data, matched experimental or quasi-experimental designs, or assessment of agreement.

One may also wonder why additional analysis strategies for comparing proportions over only a few time points are needed when more sophisticated analyses, such as growth curve models, are available. Growth curve analyses are valuable for analyzing longitudinal data sets with at least several time points, and analysis of change with two or only a few time points has limitations (Rogosa, 1980; Rogosa & Willett, 1983); but data sets with additional time points are expensive and time consuming to collect and may not be available in many cases (Henk & Castro-Schilo, 2016). This review points to additional analysis approaches that can be valuable for analyzing longitudinal studies in which only a few waves are available, in particular, when there is an interest in describing proportions at each time point and making comparisons among them. Although there is no statistical distinction among the several specification approaches that might be used, such as the latent change model (McArdle, 2001) or gain score analysis (Rovine & Molenaar, 2003), the contrast coding model has a familiar form based on latent growth curve models and a specification approach with convenient properties for specifying a variety of desired contrasts. Regardless of the specification approach, these simple structural equation models can be incorporated into

larger models to investigate more elaborate research questions, including mixed between-subjects and within-subjects interactions, proportional differences in latent class membership, covariate adjustments, paired contrasts among several repeated measures, and mediational models.

Using structural equation modeling software for categorical analysis for repeated measures or matched pairs unlocks the potential for applying state of the art estimation methods, such as diagonal weighted least squares or Bayesian estimation, and robust standard error approaches to address distributional concerns or model estimation difficulties, adding dimension of flexibility that is not available with conventional categorical data analysis methods. The focus of this paper has been on binary maximum likelihood estimation to facilitate comparison to conventional tests, but repeated measures analysis of ordinal or multinomial data is another possible application. Because structural equation modeling software conveniently handles missing data, an important advantage of comparing matched proportions with this approach is that full maximum likelihood estimation is easily implemented in the to handle missing data. Missing data are pervasive in social science research and, except for rare examples, researchers applying categorical data analysis commonly handle missing data using listwise deletion, a non-optimal approach (e.g., Enders, 2001). Moreover, inclusion of auxiliary variables, also conveniently accomplished within SEM software programs, may increase the chances of meeting missing data estimation assumptions (Collins, Schafer, & Kam, 2001).

This consideration of simple models for repeated measures and matched pairs provides insight into the fundamental concepts behind more complex longitudinal models but also illuminates a new set of tools for transcending current approaches to categorical analysis.

References

- Agresti A (2013). Categorical data analysis (Third Edition). New York: John Wiley & Sons.
- Bock RD, & Aitkin M (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, 443–459.
- Christofferson A (1975). Factor analysis of dichotomized variables. *Psychometrika*, 40, 5–32.
- Collins LM, & Lanza ST (2010). Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences (Vol. 718). John Wiley & Sons.
- Collins LM, Schafer JL, & Kam CM (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological methods*, 6(4), 330–351 [PubMed: 11778676]
- Cox DR (1958). Two further applications of a model for binary regression. *Biometrika*, 45, 562–565.
- Enders CK (2001). The performance of the full information maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement*, 61, 713–740.
- Goodman LA (1974). The analysis of systems of qualitative variables when some of the variables are unobservable. Part I. A modified latent structure approach. *American Journal of Sociology*, 79, 1179–1259.
- Henk CM, & Castro-Schilo L (2016). Preliminary Detection of Relations Among Dynamic Processes With Two-Occasion Data. *Structural Equation Modeling: A Multidisciplinary Journal*, 23, 180–193.
- Klingenberg B, & Agresti A (2006). Multivariate extensions of McNemar's test. *Biometrics*, 62, 921–928. [PubMed: 16984337]
- Lazarsfeld PF, & Henry NW (1968). Latent structure analysis. Boston: Houghton Mifflin.
- Lo Y, Mendell N, & Rubin D (2001). Testing the number of components in a normal mixture. *Biometrika*, 88, 767–778.

- Lubke GH, & Muthén B (2005). Investigating population heterogeneity with factor mixture models. *Psychological methods*, 10, 21–39. [PubMed: 15810867]
- MacCallum RC, Zhang S, Preacher KJ, & Rucker DD (2002). On the practice of dichotomization of quantitative variables. *Psychological methods*, 7, 19–40. [PubMed: 11928888]
- Masyn KE, Petras H, & Liu W (2014). Growth Curve Models with Categorical Outcomes In *Encyclopedia of Criminology and Criminal Justice* (pp. 2013–2025). Springer New York.
- Masyn KE, Petras H, & Liu W (2014). Growth Curve Models with Categorical Outcomes In *Encyclopedia of Criminology and Criminal Justice* (pp. 2013–2025). Springer New York.
- MacCallum RX, Zhang S, Preacher KJ, & Rucker DD (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7, 1–40.
- McArdle JJ (2001). A latent difference score approach to longitudinal dynamic analysis In Cudeck R, Du Toit S, & Sörbom D (Eds.), *Structural equation modeling: Present and future* (pp. 341–380). Lincolnwood, IL: Scientific Software International.
- McNemar Q (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12, 153–157. [PubMed: 20254758]
- Mulaik SA (2009). *Linear causal modeling with structural equations*. Boca Raton, FL: Chapman & Hall/CRC.
- Muthén B (1993). Goodness of fit with categorical and other non-normal variables In Bollen KA, & Long JS (Eds.), *Testing Structural Equation Models* (pp. 205–243). Newbury Park, CA: Sage.
- Muthén B (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, 43, 551–560.
- Muthén B (1996). Growth modeling with binary responses In Eye AV, & Clogg C (Eds.), *Categorical Variables in Developmental Research: Methods of Analysis* (pp. 37–54). San Diego, CA: Academic Press.
- Muthén B (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class-latent growth modeling In Collins LM & Sayer AG (Eds.), *New methods for the analysis of change* (pp.291–322) Washington, DC: American Psychological Association.
- Muthén B (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling In Collins LM & Sayer A (Eds.), *New Methods for the analysis of change* (pp. 291–322). Washington, D.C.: APA.
- Muthen B, & Christoffersson A (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika*, 46, 407–419.
- Muthén BO, du Toit S, & Spisic D (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Retrieved from http://www.statmodel.com/bmuthen/articles/Article_075.pdf
- Muthén LK, & Muthén BO (1998–2012). *Mplus User's Guide*, Seventh Edition Los Angeles: Muthén & Muthén.
- Newsom JT (2002). A multilevel structural equation model for dyadic data. *Structural Equation Modeling: A Multidisciplinary Journal*, 9, 431–447.
- Newsom JT (2015). *Longitudinal Structural Equation Modeling: A Comprehensive Introduction*. New York: Routledge.
- Radloff LS (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychology and Measurement*, 1, 385–401.
- Rasch G (1961, 6). On general laws and the meaning of measurement in psychology In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability* (Vol. 4, pp. 321–333). Berkeley, CA: University of California Press.
- Rogosa DR (1980). A critique of cross-lagged correlation. *Psychological Bulletin*, 88, 245–258.
- Rogosa DR & Willett JB (1983). Demonstrating the reliability of the difference score in the measurement of change. *Journal of Educational Measurement*, 20, 335–343.

- Rovine MJ, & Molenaar PCM (2003). Estimating analysis of variance models as structural equation models In Pugesek B, Tomer A, & von Eye A (Eds.), *Structural equation modeling: Applications in ecological and evolutionary biology research* (235–280). New York: Cambridge.
- Santor DA & Coyne JC (1997). Shortening the CES-D to improve its ability to detect cases of depression. *Psychological Assessment*, 9, 233–243
- Satorra A (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika*, 54, 131–151.
- Sorkin DH, & Rook KS (2004). Interpersonal control strivings and vulnerability to negative social exchanges in later life. *Psychology and Aging*, 19, 555–564. [PubMed: 15584782]
- Stuart A (1955). A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika*, 42, 412–416.
- Voelkle MC (2008). Reconsidering the use of autoregressive latent trajectory (ALT) models. *Multivariate Behavioral Research*, 43, 564–591. [PubMed: 26771428]

(a)

(b)

(c)

(d)

Figure 1.

Four structural equation modeling specifications equivalent to a repeated measures ANOVA: (a) a single gain score, (b) two correlated measured variables, (c) simple latent growth curve (contrast coding) model, and (d) simple latent difference score model.

Figure 2.
Contrast coding model for comparing two related proportions.

Figure 3.
Contrast coding model for comparing three related proportions.

Figure 4.
Second-order contrast coding model with two latent class classes

Table 1.

Frequency (percent) of respondents exceeding the clinical depression cutoff at Time 1 and Time 2

Time 1	Time 2		
	<u>Non-depressed</u>	<u>Depressed</u>	<u>Total</u>
Non-depressed	163 (22.67)	173 (24.06)	336 (46.73)
Depressed	51 (7.09)	332 (46.18)	383 (53.27)
Total	214 (29.76)	505 (70.24)	719 (100.00)