Published in final edited form as:
Multivariate Behav Res. 2013 January 1; 48(1): 1–27. doi:10.1080/00273171.2012.730072.

A Comparison of Factor Score Estimation Methods in the Presence of Missing Data: Reliability and an Application to Nicotine Dependence

Ryne Estabrook and **Michael Neale**
Virginia Commonwealth University

Abstract

Factor score estimation is a controversial topic in psychometrics, and the estimation of factor scores from exploratory factor models has historically received a great deal of attention. However, both confirmatory factor models and the existence of missing data have generally been ignored in this debate. This article presents a simulation study that compares the reliability of sum scores, regression-based and expected posterior methods for factor score estimation for confirmatory factor models in the presence of missing data. Although all methods perform reasonably well with complete data, expected posterior-weighted (full) maximum likelihood methods are significantly more reliable than sum scores and regression estimators in the presence of missing data. Factor score reliability for complete data can be predicted by Guttman's 1955 formula for factor communality. Furthermore, factor score reliability for incomplete data can be reasonably approximated by communality raised to the $\frac{1}{1-P(Missing)}$ power. An empirical demonstration shows that the full maximum likelihood method best preserves the relationship between nicotine dependence and a genetic predictor under missing data. Implications and recommendations for applied research are discussed.

Exploratory and confirmatory factor analyses are common tools in psychological research; they enable the assessment of latent constructs that are otherwise difficult to study and analyze. Whereas factor analysis is a method for describing samples or populations, the results of a factor analysis may be used to generate factor estimates or predictions for each individual in a sample, called *factor scores*. These factor scores may be used in a variety of ways, including distributional checks on factor models, use as individual measures in clinical and applied settings, and inclusion in other types of analyses.

Despite their utility, factor scores and their estimation are subject to a number of problems, many of which have been discussed in detail elsewhere (see Bartholomew, Deary, & Lawn, 2009; Mulaik, 2005, for recent reviews). The majority of these problems can be described as some form of indeterminacy. Wilson (1928) pointed out the indeterminacy inherent to the factor model as the number of common and unique latent factors in any model exceeds the number of observed variables. Guttman (1955) quantified the effects of indeterminacy in terms of factor communalities (ρ^2), or the squared multiple correlation between a factor and the set of manifest variables. Furthermore, some factor score estimation methods can yield factor means and variances different from those found in the original factor model (Lawley & Maxwell, 1971).

The characteristics of factor score methods for exploratory factor analysis have been studied previously (Acito & Anderson, 1986; Grice, 2001a; 2001b; Tucker, 1971). Exploratory and

Correspondence concerning this article should be addressed to Ryne Estabrook, Virginia Commonwealth University, VIPBG, P.O. Box 980126, Richmond, VA 23298. crestabrook@vcu.edu.

confirmatory factor analyses are closely related, but they differ in their methods of estimation. Of particular note is that confirmatory factor models can be applied to raw data, categorical variables, and missing values. Confirmatory factor analytic methods can and should take advantage of the tools used in exploratory factor analysis, but the characteristics of factor score estimation should be reevaluated when applied to new estimation procedures and more complex types of data.

The focus of this article is to evaluate several methods for factor score estimation for confirmatory factor models when applied to data sets with missing values. We first discuss the factor score estimation methods to be compared, then evaluate the reliability of each of these methods for continuous data, and finally evaluate the impact of different types of missing data values on the estimators.

FACTOR SCORE ESTIMATION METHODS

There exist multiple methods for the estimation of factor scores, and they vary in their complexity and application. In this section, we briefly review three of the more commonly used approaches to scoring: regression-based or exact methods, sum scores and approximate methods, and expected posterior methods.

Regression or Exact Methods

Regression or exact factor score methods use the estimated parameters from a factor analysis to define linear combinations of observed variables that generate factor scores. These methods are among the most widely used, encompassing a number of related methods for defining the necessary regression equations. Thomson's method (Thomson, 1935), which is sometimes cited as Thurstones's regression method (Thurstone, 1935), defines the factor score as the product of three terms: the factor loading matrix (Λ), the inverse of the data covariance matrix (Σ^{-1}), and the data vector of interest (y_i), as shown in Equation 1. An alternative specification of this method is given in Equation 2.

$$\widehat{f}_{i|Thomson} = \Lambda' \sum^{-1} y_i \quad (1)$$

$$= (I + \Lambda' \Theta^{-1} \Lambda)^{-1} \Lambda' \Theta^{-1} y_i \quad (2)$$

Bartlett's method (Barlett, 1937; see Lawley & Maxwell, 1971, Chapter 8) is a correction to Thomson's (1935) method that corrects for bias in factor means. Given a factor loading matrix Λ, a factor covariance matrix Φ, and a residual covariance matrix Ψ, Bartlett's method is given in Equation 3.

$$\widehat{f}_{i|Barlett} = (\Lambda' \Phi^{-1} \Lambda)^{-1} \Lambda' \Psi^{-1} y_i \quad (3)$$

Additional regression-based estimators include Anderson and Rubin's (1956) extension of Bartlett's (1935) estimator for orthogonal factors; McDonald's (1981) and the Ten Berge, Krijnen, Wansbeek, and Shapiro (1999) oblique versions of this extension; Harman's (1976) "idealized variable" estimator; and the least-squares solution described by Horst (1965). These estimators and their characteristics are described in greater detail by Grice (2001b).

Approximate Methods

Alternatives to the aforementioned regression-based methods rely not on the exact values of the estimated model parameters but rather approximations of them. Horn, McArdle, and Mason (1983) proposed a simplification of both factor analysis and factor score estimation procedure that restricts the possible values for factor loadings and factor score regression weights. For the purposes of factor score estimation, a criterion value is established where loadings more extreme than the criterion are considered "salient" and assigned a value of ±1, whereas all other loadings are assigned a value of 0. This defines the factor score estimator as the sum of the items with salient factor loadings, as given in Equation 4.

$$\widehat{f_i} = \sum_{i=1}^{k} W_i y_i, \quad W_i \in [-1, 0, 1] \quad (4)$$

Factor scores generated by this method are more commonly referred to as "sum scores"—Equation 4 describes a simple sum of salient items. Sum scores are often used in other contexts with or without a measurement model, but alternative unit weighting procedures have been in use in test battery construction (integral gross score weighting; Wherry & Gaylord, 1946) and as a sufficient statistic for trait scores in the Rasch model (Rasch, 1960) prior to Horn et al.'s (1983) work.

Items should be factor analyzed to check for reverse scoring and nonsalient loadings prior to treating sum scores as factor score estimates.

Expected Posterior Methods

Expected posterior or expected a posteriori (EAP) methods apply Bayes' theorem to the estimation of randomly varying latent traits in one of two ways. Factor scores and other random effects may be estimated using Markov chain Monte Carlo (MCMC) procedures like any other Bayesian optimization, with factor scores taken as the mean of each individual's posterior distribution. Alternatively, expected posterior methods may be applied as the second step of a two-stage analysis, where EAP methods are applied using the results of a previously fitted model. The probability of any value for an individually varying latent trait f_i is given in Equation 5.

$$P(\widehat{f_i}|y_i) = \frac{P(\widehat{f_i})P(y_i|\widehat{f_i})}{\int P(\widehat{f_i})P(y_i|\widehat{f_i})} \quad (5)$$

Statistical software packages for structural equation modeling typically employ regression-based techniques rather than EAP methods for continuous data.[1] Expected posterior methods are more common in item response models, which often use EAP as a method for generating trait scores.

[1]AMOS supplies regression weights for manual factor score calculation (Arbuckle, 2006); EQS applies regression and Bentler-Yuan methods (Bentler, 1995); LISREL uses the Anderson-Rubin method (Jöreskog & Sörbom, 2006); Mplus applies the Bartlett method for continuous data and weighted least squares (WLS) methods and EAP for maximum likelihood models using categorical or count data (Muthén & Muthén, 1998–2007).

MAXIMUM LIKELIHOOD METHOD FOR EXPECTED POSTERIOR FACTOR SCORES

An alternative to the aforementioned methods is to estimate factor scores as free parameters. Under this method, the estimates from a factor model fit to a sample are used to create a model for each individual that includes factor scores as the only free parameters.

Model Estimation

Just as with the factor score methods described previously, maximum likelihood factor score estimation depends on results from a previously fitted factor analytic model. The estimates from that model are used to populate a model fit to a single individual or row of data. All factor loadings (Λ), item intercepts (α), and residual variance and covariance structures (Θ) are fixed at their estimated values.

$$Mean=\alpha+\widehat{f_i}\Lambda \quad (6)$$

$$Cov=\Theta \quad (7)$$

Likelihood Function

A common critique of maximum likelihood estimation is that it yields not the likelihood of the parameters given the data but rather the likelihood of the data given the estimated parameters. When we restrict our discussion to the estimation of factor scores, simply including factor scores as free parameters in a model would estimate the likelihood of the data given the current factor score estimate, $L(y_i \mid \hat{f}_i)$, whereas the likelihood of the factor score estimate given the data,.$(\hat{f}_i \mid y_i)$, is actually of interest. This can be resolved by applying the expected posterior method described earlier to the likelihood function used in our factor score model. The application of Bayes' theorum given in Equation 8 shows that generating the desired likelihood would require knowledge of the likelihood of the factor score estimate, $L(\hat{f}_i)$.

$$L(\widehat{f_i}|y_i)=L(\widehat{f_i})L(y_i|\widehat{f_i}) \quad (8)$$

The estimated factor means and covariances can be used to define the likelihood of a given set of data drawn from a particular normal distribution. If we define the likelihood of a particular factor score estimate $\hat{f}_1$ using the density of the multivariate normal distribution (ϕ) as shown in Equation 9, that density may be substituted and yield the model likelihood shown in Equation 10.

$$L(\widehat{f_i})=\varphi(\widehat{f_i}|\mu_f,\Phi) \quad (9)$$

$$L(\widehat{f_i}|y_i)=\Phi(\widehat{f_i}|\mu_f,\Phi)L(y_i|\widehat{f_i}) \quad (10)$$

This new likelihood is simply the product of the usual (albeit conditional) SEM maximum likelihood function and a scalar weight defined by the multivariate normal density function. Specifying a model for these factor scores requires either replacing a standard maximum likelihood estimator with the full maximum likelihood function in Equation 10 or weighting

the standard estimator by the density function found in Equation 9. A guide to this specification in OpenMx (Boker et al., 2011) is shown in Appendix A.

CURRENT STUDY

This study is a comparison of the various two-stage factor score estimators described earlier using a Monte Carlo simulation. There are two purposes to this study. One purpose is evaluate the relative performance of the tested factor score estimators to determine which method or methods perform best and to what extent those differences depend on missing data and other data features like sample size, number of items, and the magnitude of factor loadings. Another purpose is to evaluate the absolute rather than relative performance of the listed factor score estimators to establish whether and when factor score estimates are reliable measures of latent traits. By comparing this set of factor score estimations over varying missing data patterns and other data characteristics, we can determine the appropriateness of factor scoring methods in a wide range of possible empirical situations and do so with far greater control than we would have with empirical data.

SIMULATION

The following simulation is designed to evaluate the reliability of factor score estimation methods for continuous data. A total of four methods were compared: sum scores, the Bartlett or regression-based scores, a full maximum likelihood (ML) estimator that uses the expected posterior weighting approach and the likelihood function defined in Equation 10, and an un weighted ML estimator that uses the standard maximum likelihood estimator without expected posterior weighting. First, data were generated under an assumed common factor model varying the number of observed variables, loadings, and other relevant data parameters. A confirmatory factor model was then fit to the generated data, and factor scores were generated under each method. Data generation, model fitting, and factor score estimation were repeated on 10 data sets per simulation condition. Finally, the reliabilities of each method were computed and compared. Each of the aforementioned steps is described in greater detail in the ensuing sections.

Data Generation

Data were generated under the common factor model shown in Equation 11, where Y is a matrix of observed data with n individuals and k variables, Ξ is an $n \times 1$ matrix of individually varying common factor scores, Λ is a $1 \times k$ matrix of factor loadings, E is an $n \times k$ matrix of standardized residuals, and Θ is a diagonal matrix of residual standard deviations.

$$Y_{n\times k}=\Xi_{n\times 1}\Lambda_{1\times k}+E_{n\times k}\Theta_{k\times k} \quad (11)$$

First, individual factor and error scores were randomly drawn from standard normal distributions using the R function rnorm (), and Λ and Θ were generated based on the simulation condition. Once generated, the matrices Ξ, Λ, E, and Θ were used to generate the observed data Y using Equation 11. In conditions where missing data were required, the data for any person i and variable j were assigned as missing following a random draw from a binomial distribution with probability p; observations selected by the binomial draws were deleted from the data to create missing values. The number of items, factor loadings, sample size, and missing data pattern (if present) were varied across conditions as described here and summarized in Table 1.

Items—Data were generated with either 3, 5, 10, or 20 observed continuous variables.

Factor loadings (communalities)—The magnitude of the factor loadings used for data generation were varied across simulation conditions. The mean factor loading across all items was fixed to values of 0.5, 0.7, or 0.9. These values correspond to item communalities (squared correlations between items and the common factor, ρ) of 0.25, 0.49, and 0.81 for each condition, indicating that 25%, 49%, and 81% of the observed variance in y_i is attributable to the common factor in each condition. All factor loadings were generated as standardized loadings such that the communality and residual variance for any variable always sum to a value of one.

The loadings also varied across items within each condition. Variance in factor loadings was created by drawing communalities from a normal distribution with a given mean (0.25, 0.49, or 0.81, as described earlier) and standard deviation (0.0, 0.2, or 0.4). The quantiles of the assigned distribution of communalities were drawn[2] and then adjusted to be between 0.05 and 0.95 to ensure that no item either provided no information or was too perfect an indicator of the common factor.[3] The square root of these communalities were then used as factor loadings. For example, the simulation condition with three items, a mean factor loading of 0.7, and a standard deviation of 0.2 would draw the 25th, 50th, and 75th quantiles from a distribution of communalities ($\rho^2 \sim N.(0{:}49;\ 0{:}2)$. The resulting communalities of 0.355, 0.490, and 0.625 would respectively yield factor loadings of 0.596, 0.700, and 0.791 and residual variances of 0.645, 0.510, and 0.375 for the three items.

Sample size—Each iteration of the simulation had a sample size of 100, 200, or 500.

Missing data—This simulation study contained both full data and missing data conditions. When full data were requested (i.e., no missing data were generated), individually varying factor and error scores were generated as described previously and used to generate observed data given the simulation conditions described earlier. To create missing data, full data were first created and then deleted according to one of two missing data mechanisms. Base probabilities of any observation being missing were assigned values of 0.05, 0.10, 0.15, 0.20, 0.25, or 0.30 for any condition.

Two types of missing data mechanisms were used in this simulation. In the first condition, data were simulated to be missing completely at random (MCAR; Rubin, 1976). Full data were generated, then individual observations were deleted for each variable, with the probability of deletion/missingness equal to the missingness variable for that simulation condition. This missing data mechanism is expressed in Equation 12, with y_{ij} representing an observed score for person i on variable j, R_{ij} a binary variable that indicates whether y_{ij} is observed or missing, and δ representing the assigned missingness probability for a given simulation condition.

$$P(R_{ij}|y_i, MCAR)=\delta \quad (12)$$

In the second condition, data were simulated to be missing at random (MAR), indicating that missingness depends on other observed data values. The probability of missingness for any person and item was generated as the assigned missingness probability for a given simulation condition δ plus variable describing the impact of the other observed variables z_{ij}. This variable

[2]For k items, the $\frac{1}{k+1}$ th through $\frac{k}{k+1}$ th quantiles of the given normal distribution were drawn as communalities. Using quantiles rather than random draws ensured a symmetric distribution of communalities with a mean equal to the assigned condition mean.

[3]Communalities less than 0.05 were fixed to be 0.05 and communalities greater than 0.95 were fixed to be 0.95. Whenever the $\frac{i}{k+1}$ th quantile was adjusted up or down to meet this range restriction, the $\frac{k-i}{k+1}$ th was adjusted by the same amount in the opposite direction to ensure that the communalities were fixed to have the same mean as assigned by the simulation design. For example, the three-item condition with a mean loading of 0.9 and variance of 0.4 yields communalities of 0.54, 0.81, and 1.08. In this condition, the third communality would be adjusted down by 0.13 to 0.95, and the first communality would be adjusted up 0.13 to 0.67 to preserve a mean value of 0.81 at the expense of reduced variance.

was defined as a function of the observed data any individual, such that z_{ij} is a function of the first through.$(j-1)^{th}$ items of y_i. This variable was constrained to have a mean of 0 and fall in the range [−0.05, 0.05] to maintain a constant effect across all simulation conditions and not create negative missingness probabilities in conditions where δ = 0:05. A brief description of this missing data mechanism is given in Equations 13–15; greater detail regrading the MAR data generation used in this simulation is available in Appendix B.

$$P(R_{ij}|y_i, MAR)=\delta+z_{ij} \quad (13)$$

$$z_{ij}=f(y_{i,1:(j-1)}) \quad (14)$$

$$z_{ij} \in [-0.05, 0.05] \quad (15)$$

Model fitting—Once generated, the data were fit with a common factor model using full information maximum likelihood (FIML) in OpenMx (Boker et al., 2011). To identify the model, the common factor mean and variance were set to zero and one, respectively, allowing all factor loadings, residual (unique) variances, and item intercepts to be freely estimated. This led to the model expected means and covariances given in Equations 16 and 17, where α is a vector of manifest intercepts, μ is a vector of factor means, Λ is a matrix of factor loadings, Φ is a factor covariance matrix, and Θ is a diagonal matrix of residual variance terms. For this single factor model, μ and Θ were fixed to zero and one, respectively.

$$Mean=\alpha+\mu\Lambda \quad (16)$$

$$Cov=\Lambda'\Phi\Lambda+\Theta \quad (17)$$

The parameters from converged models were extracted and used for factor score estimation in the next step of the simulation.

Factor Score Estimation

Four factor scoring methods were calculated or estimated. Sum scores were calculated not with summation but rather as each person's individual mean across items times the number of items. This method is equal to a simple sum when no missing data are present but keeps the scale of the sum scores constant across all patterns of missingness. Bartlett scores were calculated using the definition given in Equation 3. When missing values were present for any individual, only the matrix elements corresponding to the nonmissing variables were used.

Both full and un weighted versions of maximum likelihood factor scores were estimated using separate OpenMx models for each individual. Regardless of the presence of a weight, each individual's factor score was estimated from a model that included that individual's row of data (with missing values where appropriate) as the entire data set for that model. All factor loadings, residual variances, and manifest intercepts for that model were fixed at the values found in the initial model estimation. The only free parameter in both ML methods was the factor score estimate. Both methods estimated models with the expected means and covariances given in Equations 6 and 7, indicating that the expected mean of the observed data was dependent on the estimated factor score ($\xi_i\Lambda$) and the expected covariance conditional on the means is the estimated residual variances Θ. When missing values were present, the FIML method proceeds by eliminating those rows and columns of the predicted covariance matrix

corresponding to the position of the missing values. Elements of the predicted mean vector (or threshold matrix) are also removed, so the likelihood calculation is performed only for those values present in the data.

Comparing Factor Score Methods

The primary metric for comparing the performance of the factor score methods in this study is reliability, defined in terms of the *reliability index*. The reliability index is the correlation between an observed score and its underlying true score. Although the reliability index cannot be calculated on empirical data as true scores are not known, this measure is easily calculated in simulation studies. Squaring the reliability index yields the proportion of observed score variance attributable to true score variance, one of the most common definitions of reliability and the estimate of reliability used in this study.

RESULTS

Mean factor score reliability by number of items and missingness mechanism is presented in Table 2. The full maximum likelihood estimator outperforms both sum scores (t_{107} = 7:38, p <:001) and the Bartlett method (t_{107} = 4:08; p <:001) while showing no improvement over the unweighted ML estimator (t_{107} = −1:00; p =:319). When missing data are present, full ML maintains its edge over sum scores (MCAR: t_{647} = 14:52; p <:001; MAR: t_{647} = 13:95; p <: 001) and the Bartlett method (MCAR: t_{647} = 19:04; p <:001; MAR: t_{647} = 19:24; p <:001) regardless of missing data mechanism. However, full ML estimates show an advantage over the unweighted estimates when missing data are present (MCAR: t_{647} = 11:34; p <:001; MAR: t_{647} = 11:46; p <:001), indicating the importance of the empirical Bayes component of maximum likelihood factor score estimation.

Predicting Reliability From Simulation Parameters

Although the comparison of the various factor score methods is important, it is also important to identify the major predictors of measurement precision. The effects of the simulation parameters on mean full ML factor score reliability for each complete data condition are shown in Table 3. These results show strong effects of both the number of items and their overall magnitude and a smaller effect of the variance in loadings.

Guttman 1955 has shown that factor score reliability should be a function of factor communality, or the squared multiple correlation between manifest variables and a factor (ρ^2). Factor score communality is calculated as the quadratic product of the factor loading matrix Λ and the inverse data covariance matrix Σ ($\rho^2 = \Lambda\Sigma^{-1}\Lambda'$). Regressing mean factor score reliability on communality yields a near-perfect relationship, as shown by the R^2 values in Table 4 (R^2 = 0:996). However, the slope of this line is not the 1:1 relationship predicted by Guttman, instead showing that observed factor score reliabilities are slightly lower than the estimated communalities for each condition.

A comparison of communalities and estimated factor score reliabilities is presented in Figure 1. Although the regression presented in Table 4 (regression line in solid black) provides the best fit, the residuals in this model are highly negatively skewed (Full Sample: σ^3 = −4:72; Condition Means Only: σ^3 = −1:16). Further analysis of the plot in Figure 1 shows a number of outlying cases in the lower right portion of the graph, which both drive the regression line and predominantly consist of conditions with three items and/or sample sizes of 100, which are both prone to increased sampling variation and lower likelihood of convergence. The inability to replicate Guttman's (1955) 1:1 relationship between communality and factor score reliability may be due to poor correspondence of observed data to the factor scores used for data generation, whether due to sampling variation, poor convergence, or other factors.

The complete data findings shown in Figure 1 can also be analyzed for the missing data conditions. Figure 2 presents results for the 0%, 15%, and 30% missingness conditions for both MCAR and MAR missing data mechanisms. These graphs show greater deviations between the best-fit curve (solid) and the unit slope line (dotted) as the amount of missing data increases. Whereas the complete data conditions were best fit by a simple linear trend of communality (β*Communality), missing data conditions were best fit by a exponential relationship between reliability and communality.

The exponents for each level of missingness under MAR assumptions are shown in Figure 3. The exponents described by this figure come from regression models where log-transformed reliability is predicted by log-transformed communality. This log transformation was done to express the nonlinear relationship between communality and reliability in linear terms. Equation 18 shows the exponential relationship between the two terms, whereas Equation 19 shows that a simple linear regression can be used to express this relationship in log-transformed variables. As such, log-transformed versions of both variables were used to fit this regression equation. The inverse of these exponents ($\frac{1}{\beta}$) for each level of missingness under MAR assumptions is shown in Figure 3, which is shown in Equation 20.

$$Reliability = Communality^{\beta} \quad (18)$$

$$\log(Reliability) = \beta \log(Communality) \quad (19)$$

$$Communality = Reliability^{\frac{1}{\beta}} \quad (20)$$

There is a strong linear relationship between the amount of missingness and the exponent $\frac{1}{\beta}$ in Figure 3. The slope of these lines (B = −1:012, SE = 0:040 for all MAR cases; B = −1:034, SE = 0:009 for selected MAR cases; B = −0:963, SE = 0:036 for all MCAR cases; B = −1:038, SE = 0:062 for selected MCAR cases) closely mirrors the proportion of complete data (1-P (Miss)). Thus the reliability of factor scores under missing data appears to be reasonably approximated by factor communality raised to the k^{th} power, where $k = \frac{1}{1-P(Miss)}$ is the inverse of the proportion of non missing values in the data.

EMPIRICAL EXAMPLE

The utility of these factor scoring techniques is illustrated using data from the Virginia Adult Twin Study of Psychiatric and Substance Use Disorders (VATSP-SUD). Previous analyses of the VATSPSUD have been used to investigate genetic causes of nicotine dependence and have replicated previous work showing effects of the CHRNA3 and CHRNA5 receptor genes on smoking behavior (Chen et al., 2009; Maes et al., 2011; Maes et al., 2004). Further descriptions of this data set can be found in previous works (Maes et al., 2011). This example recreates a version of that analysis that utilizes the presented factor scoring techniques to test the effect of a single gene on nicotine dependence.

The sample for this analysis consists of the 2,610 individuals from the VATSPSUD that provided both nicotine dependence and genetic information. Nicotine dependence was measured using a modified version of the Fagerström Tolerance Questionnaire (FTQ; Fagerström, 1978). A single item (“Are you a current smoker?”) was added to the eight smoking behavior items already present in the FTQ for this example. All individuals used in this analysis were also genotyped, and the rs16969968 single-nucleotide polymorphism (SNP) from the CHRNA receptor gene cluster is used as a predictor of nicotine dependence in the ensuing

analysis. This binary genomic variable serves both to demonstrate secondary analysis as one use of factor scores and to provide a measure of factor score validity in the absence of the known true scores used to calculate reliability in the previous simulations.

First, a one-factor model was fit to the nine variable data. This model showed acceptable fit (CFI = 0.918, RMSEA = 0.061), and the nine items yielded a factor communality of 0.804. A subsequent structural equation model regressed the common nicotine dependence factor on the CHRNA variable, finding a significant association between the SNP and the common factor (B = 0:090 (.033), β=:060 (.022), χ^2_1=7.460, p =:006). This regression was then carried out using the four factor-scoring methods tested earlier, and these results are shown in standardized form on the left of Table 5. All four methods showed some shrinkage of the genetic effect, with the ML methods performing best and sum scores performing worst.

Missing data were artificially inserted into the VATSPSUD data set using the MAR procedure described in the simulation study. Twenty-five percent of responses were deleted, leaving individuals with between two and nine responses to the nicotine dependence scale. The factor models, factor scoring procedures, and regressions were then repeated for this data set, the results of which are shown in the right side of Table 5. The full ML method finds the strongest evidence of the SNP-nicotine effect both with respect to the size of the standardized effect and its associated p value.

DISCUSSION

This article presents a simulation study comparing four factor score estimators when individual responses may include missing values. We found that estimating factor scores by maximum likelihood outperformed traditional factor score predictors. Although the differences between the four estimators were relatively small with complete data, the relative superiority of full maximum likelihood estimation increased when missing values were present in the data. The superiority of weighted maximum likelihood methods was further shown in the empirical example, as this method best preserved the effect found in a single-step structural equation model.

The performance of each of the tested estimators under missing data can be directly mapped to their method of missing data handling. The Bartlett approach does not provide any method for handling missing data, so cases with any missing values are simply omitted. For centered data, this amounts to imputing missing values with the population mean (zero). Sum scores perform better than the Bartlett method in this simulation because sum scores were defined not as a simple sum but rather a rescaling of the individual mean.[4] This mechanism is equivalent to imputing missing values with each individual's mean rather than the population mean, which gives sum scores their advantage over Bartlett scores. The final two estimators utilize full-information maximum likelihood, which provides further robustness to missing data by avoiding imputation altogether in favor of defining the expected factor score given the available data. As the likelihood function in the full ML version of this estimator better describes this expectation, it performed better than unweighted maximum likelihood when missing data were present. Just as full-information methods provide advantages over multiple imputation in structural equation modeling, full-information factor score estimators provide advantages over imputation-based factor score methods.

Although the full maximum likelihood performs better overall, it does so at the expense of added computational demand. Both sum and Bartlett scores are easy to calculate and require

[4]Alternatively, this can be viewed as a special case of complete case analysis, where the completed items are considered representative of the missing items.

trivial computation time. Given the small performance differences between the Bartlett estimator and the full ML method for complete data, there is little to be gained in terms of precision from using full maximum likelihood for complete data with no missing values. However, full ML should be used when factor scores are to be extracted from data sets that include missing values. Regardless of the method used, factor scores for a given data set and analysis should all be estimated using the same method to avoid potential scaling differences between factor score estimators.

One unexplored advantage of full-information methods for factor scoring relates to individual differences in precision. As an individual's amount of missing data increases, the precision with which researchers can estimate their individual factor scores should decrease. The utility of estimated standard errors, confidence intervals, and other measures of factor score precision are an important topic for future study, though it should be noted that the method described here does not account for the imprecision of factor loadings, residual variances, and other estimated parameters and thus overstates the precision by which factor scores are measured. Previous research has shown that individual likelihoods can be used as a measure of how consistent an individual's responses are with the population model (Neale, 2000; Reise & Widaman, 1999). Given sufficient sample size, it is possible to compare between individuals with the same pattern of missingness. If the items are effectively interchangeable (i.e., having equivalent response frequencies and factor loadings) then comparing between individuals with the same number of missing items is a practical alternative.

Although it is possible to compute an approximate *z* score that corrects for the number of missing items (Hopper & Mathews, 1982), the statistical properties of these measures of fit have not been fully explored and may be in need of refinement and testing, especially in a psychological measurement context. Researchers interested in more exact measures of precision may look to one-step methods (including Bayesian estimation of factor models and factor scores) to better model the impact of parameter precision. Properly accounting for individual differences in missingness patterns and related differences in precision are crucial to using missingness-dependent factor scores in secondary analyses.

Although this study largely replicated Guttman's (1955) relationship between communality and factor score reliability, the lack of a perfect linear relationship between factor communality and the reliability of full ML factor scores merits further analysis. There are several possible explanations. As noted in the results, the skewness of the regression residuals and further graphical analysis of the results in Figure 1 show that the simulation includes a number of outliers, particularly in simulation conditions with small samples and three items. Another possibility is that the metric of reliability used, namely, the reliability index, may differ from empirical reliability measures. The two-step factor score procedure allows for two sources of error; differences between true and estimated factor scores may be induced not only by the indeterminacy of the model but also differences between the true and estimated factor structure. It is possible that the relationship between communality and reliability only approaches a unit relationship as sampling variation asymptotically approaches zero.

The design of our simulation may have underestimated the value of maximum likelihood in other empirical situations. In this report we generated missing values under a limited number of specific conditions, which are examples of the very general scenarios of MCAR and MAR to which maximum likelihood estimates (MLEs) are known to be robust. Furthermore, this simulation used cases where items were approximately equal with regard to their means, scale, and propensity for missingness. Additional testing is needed to compare the properties of these factor score methods in other empirical contexts, but the gap between full-information methods and other methods should only increase as the set of manifest variables becomes more heterogeneous with regard to these structural elements. Although MLEs are known to be biased

if the data are missing not at random (Rubin, 1976) it is possible to model hypothesized missingness mechanisms in these circumstances.

There are some caveats to the results presented here. One is that using a two-step procedure to estimate first the parameters of the model and second the MLEs of the factor scores using these parameter estimates is (like most two-step statistical procedures) suboptimal. Although the estimates of the factor scores are asymptotically unbiased, estimates of their standard errors or likelihood-based confidence intervals would not be. The empirical example shows some degree of shrinkage of the nicotine-gene effect in all factor scoring methods relative to the single-step SEM model. A Bayesian perspective, in which both the parameters of the model and the data are subject to error, would suggest that the two-step ML procedure would overestimate the precision of any particular factor score. Although Bayesian analysis has its advantages, it can be slow and relatively difficult to generalize to different classes of problem and is less predictable than numerical optimization using the Newton-Raphson and similar methods common in structural equation modeling. If the precision of the factor scores is of key interest, then one might use bootstrap methods to approximate the multivariate distribution of the model parameters and use each instance of the parameter estimates to estimate the factor scores. Another possibility might be to use sparse optimization methods to estimate jointly the parameters of the model and the factor scores. A second issue that should be noted is that we provided an approximation of the communality predictor of factor score reliability under missing data. It is not a precisely derived, formulaic estimate of reliability. In cases of complex types and patterns of missingness, researchers interested in factor score reliability would be better served by adapting the presented simulation and generating estimates of factor score reliability tailored to the characteristics of their particular study.

Although the results of this simulation provide an initial comparison of factor scoring methods, full maximum likelihood methods in particular must be investigated under more complex factor models and data types. As the proposed full maximum likelihood method utilizes the factor covariance matrix when calculating weights, this method should have desirable properties when applied to models with multiple factors. Using factor correlations when calculating density weights should improve recovery of factor correlations. Simultaneously estimating scores for correlated factors may improve reliability by taking advantage of the relationship between a particular factor and all items it correlates with, not just the items it directly loads on. Furthermore, the normal density weighting used in this approach could be replaced with alternate factor distributions when evidence of factor nonnormality is detected (J. E., S. H., P. D., T. S., & Neal, 2006). Whether or not full maximum likelihood factor scores actually have these properties is an empirical question and should be a topic of later studies. Similar extensions to data sets with ordinal and both ordinal and continuous measures are also important directions for future work.

The results of the simulation indicate that the full ML method produces reliable factor score estimates and that the reliability of these factor scores can be closely approximated by missingness-adjusted communality. Although other factor score methods may take less time, the full ML method provides for both the most reliable estimates and the greatest flexibility as it can be applied to arbitrary structural equation models with any amount of missing data. It may also be used in a factor mixture model context in which factor scores may be estimated for all of the components of a mixture distribution. The relationship between communality and reliability gives researchers a measure of factor score reliability that can be used both to describe the quality of factor scores and inform the design of future factor analytic studies. Having a closed-form and easily calculated measure of reliability also reinforces the difference between theoretically error-free factors and the imperfection of factor score estimates. Just as previous factor score controversies have shown that factor score indeterminacy is quantifiable

and does not obviate their utility, their reliability can also be quantified and allow researchers to make an empirical rather than ideological decision regarding their use.

References

Acito F, Anderson RD. A simulation study of factor score indeterminacy. Journal of Marketing Research. 1986; 23:111–118.

Anderson RD, Rubin H. Statistical inference in factor analysis. Proceedings of the Third Berkeley Symposium of Mathematical Statistics and Probability. 1956; 5:111–150.

Arbuckle, JL. Amos 7.0 user's guide. Chicago, IL: SPSS; 2006.

Bartholomew DJ, Deary IJ, Lawn M. The origin of factor scores: Spearman, Thomson and Bartlett. British Journal of Mathematical and Statistical Psychology. 2009; 62:569–582. [PubMed: 19321036]

Bartlett MS. The statistical conception of mental factors. British Journal of Psychology. 1937; 28:97–104.

Bentler, PM. EQS structural equations program manual. Encino, CA: Multivariate Software; 1995.

Boker S, Neale MC, Maes H, Wilde M, Spiegel M, Brick T, Fox J. OpenMx: An open source extended structural equation modeling framework. Psychometrika. 2011; 76:306–317. [PubMed: 23258944]

Chen X, Chen J, Williamson VS, An SS, Hettema JM, Aggen SH, Kendler K. Variants in nicotinic acetylcholine receptors alpha5 and alpha3 increase risks to nicotine dependence. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics. 2009; 150B:926–933.

Fagerström KO. Measuring degree of physical dependence to tobacco smoking with reference to individualization of treatment. Addiction Behavior. 1978; 11:331–334.

Grice JW. A comparison of factor scores under conditions of factor obiquity. Psychological Methods. 2001a; 6:67–83. [PubMed: 11285813]

Grice JW. Computing and evaluating factor scores. Psychological Methods. 2001b; 6:430–450. [PubMed: 11778682]

Guttman L. The determinacy of factor score matrices, with implications for five other basic problems of common-factor theory. British Journal of Statistical Psychology. 1955; 8:65–81.

Harman, HH. Modern factor analysis. 3. Chicago, IL: University of Chicago Press; 1976.

Hopper JL, Mathews JD. Extensions to multivariate normal models for pedigree analysis. Annals of Human Genetics. 1982; 46(Pt. 4):373–383. [PubMed: 6961886]

Horn JL, McArdle JJ, Mason R. When invariance is not invariant: A practical scientist's view of the ethereal concept of factorial invariancesnce. The Southern Psychologist. 1983; 1:179–188.

Horst, P. Factor analysis of data matrices. New York, NY: Holt, Rinehart & Winston; 1965.

Jöreskog, KG.; Sörbom, D. Lisrel 8.8 for Windows [Computer software]. Lincolnwood, IL: Scientific Software International, Inc; 2006.

Lawley, DN.; Maxwell, MA. Factor analysis as a statistical method. 2. London, UK: Butterworths; 1971.

Maes HH, Neale MC, Chen X, Chen J, Prescott CA, Kendler KS. A twin association study of nicotine dependence with markers in the CHRNA3 and CHRNA5 genes. Behavioral Genetics. 2011; 41(5): 680–690.

Maes HH, Sullivan PF, Bulik CM, Neale MC, Prescott CA, Eaves LJ, Kendler KS. A twin study of genetic and environmental influences on tobacco initiation, regular tobacco use and nicotine dependence. Psychological Medicine. 2004; 34:1251–1261. [PubMed: 15697051]

McDonald R. Constrained least squares estimators of oblique common factors. Psychometrika. 1981; 46:337–341. Retrieved from http://dx.doi.org/10.1007/BF02293740. 10.1007/BF02293740

Mulaik, SA. Looking back on the indeterminacy controversies in factor analysis. In: Maydeu-Olivares, A.; McArdle, JJ., editors. Contemporary psychometrics. Mahwah, NJ: Erlbaum; 2005. p. 173-206.

Muthén, L.; Muthén, B. Mplus user's guide. 5. Los Angeles, CA: Author; 1998–2007. Retrieved from http://www.statmodel.com

Neale, MC. Individual fit, heterogeneity, and missing data in multigroup sem. In: Little, TD.; Schnabel, KU.; Baumert, J., editors. Modeling longitudinal and multiple-group data: Practical issues, applied approaches, and specific examples. Hillsdale, NJ: Erlbaum; 2000. p. 249-267.

R Development Core Team. R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria: 2010. Retrieved from http://www.R-project.org

Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen, Denmark: Danmarks Paedagogiske Institut; 1960.

Reise SP, Widaman K. Assessing the fit of measurement models at the individual level: A comparison of item response theory and covariance structure approaches. Psychological Methods. 1999; 4:3–21.

Rubin D. Inference and missing data. Biometrika. 1976; 63:581–592.

Schmitt JE, Mehta PD, Aggen SH, Kubarych TS, Neale M. Semi-nonparametric methods for detecting latent non-normality: A fusion of latent trait and ordered latent class modeling. Multivariate Behavioral Research. 2006; 41:427–443.

Ten Berge J, Krijnen W, Wansbeek T, Shapiro A. Some new results on correlation-preserving factor scores prediction methods. Linear Algebra and Its Applications. 1999; 289:311–318.

Thomson GH. The definition and measurement of g (general intelligence). Journal of Educational Psychology. 1935; 26:241–262.

Thurstone, LL. The vectors of mind. Chicago, IL: University of Chicago Press; 1935.

Tucker LR. Relations of factor score estimates to their use. Psychometrika. 1971; 4:427–435.

Wherry RJ, Gaylord R. Test selection with integral gross score weights. Psychometrika. 1946; 11:173–183. [PubMed: 20288939]

Wilson E. On heirarchical correlational systems. Proceedings of the National Academy of Science. 1928; 14:283–291.

APPENDIX A

OpenMx Specification

This appendix outlines an OpenMx-based factor score estimation procedure. OpenMx (Boker et al., 2011) is a free open-source SEM program that is run as an R library (R Development Core Team, 2010). The following code and instructions will assume some familiarity with both OpenMx and R; additional demos, documentation, and user support are available at http://openmx.psyc.virginia.edu/. Although this appendix focuses exclusively on one program, the expected posterior weighting described in this article can be estimated using any structural equation or general linear model program that allows for the inclusion of either user-defined objective functions or user-defined weights.

The general structure of the coding example is as follows: First, we'll estimate a common factor model. We'll edit that factor model such that the factor mean is the only free parameter. When that model is applied to a single row of data, the factor mean will represent the factor score estimate. We'll then create a second model that takes the objective function from the first model and applies the expected posterior weighting. That second model can then be estimated for each individual for whom factor scores are desired.

Initial model specification

The following code presents the same matrix-based specification of a common factor model for five variables. The model is placed in the factorModel object and created by the mxModel function. The various data, matrices, algebras, and objective functions required to specify the model are then listed as arguments to the mxModel function. The first argument is a name for the model. This is followed by a call to the mxData function, which specifies that a previously defined data frame called example Data contains the data to be modeled and that this data frame contains raw data. The matrices α, μ, Λ,, Φ, and , are then created by the mxMatrix function calls, yielding MxMatrix objects entitled "alpha," "mu," "lambda," "phi," and "theta." The specification for the matrix "alpha" contains all named arguments for clarity (i.e., matrix type, number of rows), whereas all subsequent matrix specifications omit these for greater parsimony. The mxAlgebra function defines the expected covariance and mean algebras

presented in Equations 16 and 17, naming them “cov” and “mean.” Finally, the mxFIMLObjective specifies that full-information maximum likelihood is to be used and that the algebras “cov” and “mean” define the expected covariances and means for the columns of the data. The model is then run using the mxRun function and the results placed in an R object called factor Results.

```
# matrix specification of common factor model
factorModel <- mxModel(“Initial Model”,
mxData(exampleData, “raw”),
mxMatrix(type=“Full”, nrows=1, ncol=5,
free=TRUE, values=0, name=“alpha”),
mxMatrix(“Full”, 1, 1, FALSE, 0, name=“mu”),
mxMatrix(“Full”, 1, 5, TRUE, 0.8, name=“lambda”),
mxMatrix(“Symm”, 1, 1, FALSE, 1, name=“phi”),
mxMatrix(“Diag”, 5, 5, TRUE, 0.6, name=“theta”),
mxAlgebra(t(lambda) %*% phi %*% lambda + theta, name=“cov”),
mxAlgebra(alpha + mu %*% lambda, name=“mean”),
mxFIMLObjective(“cov”, “mean”, names(exampleData))
)
# run the model
factorResults <- mxRun(factorModel)
```

It should be noted that there are many other ways to specify the aforementioned factor model, including alternative algebras and the use of paths rather than matrices. Users should feel free to specify models in whatever style best suits their research problem, knowledge, and needs.

Factor score model

The next step is to create a model for factor scores. The object factorResults is actually an MxModel object, which contains all of the data, matrices, and algebras contained in factorModel with all free parameters updated to their estimated values as well as additional output information. This means that the factorResults object can be edited and made into a model for factor scores.

The transformation of our initial to results into a new model for factor scores as shown here. First, we copy the contents of factorResults into a new object called fsModel. Although this could be done more simply (i.e., fsModel <- factorResults), using the mxModel function allows us to easily change the name of the model.

```
# put initial results into new object and change the name.
fsModel <- mxModel(factorResults, name=“Factor Score”)
```

Next, we edit the matrices and algebras in the model. Rather than retype the entire model, we can refer to and replace some of the structures in the existing model. Named entities like matrices and algebras may be referred to by name; the $ operator is used to refer to named entities in other objects. Alternatively, slots are consistent features of specific OpenMx object types that may be referred to by the @ operator. For example, MxModel objects contain slots like data and objective, whereas MxMatrix contains slots like values and free that contain matrices of starting values and TRUE/FALSE values that indicate whether parameters are free or fixed. The following code accesses the free slots of the alpha, lambda, and theta matrices

and fixes the previously free parameters. The final line designates the factor means as free parameters as they will be used as the factor score estimates.

```
# fix free parameters at estimated values
fsModel$alpha@free[,] <- FALSE
fsModel$lambda@free[,] <- FALSE
fsModel$theta@free[,] <- FALSE
# free factor mean value
fsModel$mu@free[,] <- TRUE
```

The final changes we make to the model pertain to the objective function and data. The expected mean algebra is the same for both the original model (Equation 16) and factor score model (Equation 6), but the expected covariance changes from the full model covariance in Equation 17 to the covariance conditional on the factor, which is just the residuals (Equation 7). We can respecify the model objective by referencing our factor score model's objective slot. The option vector=TRUE is used to request that the objective function return a vector of raw likelihoods, which allows us to more easily weight the likelihood rather than the default −2 log-likelihood. Finally, we replace the full data set with the data for a single individual, making this model a factor score model for a single individual. Running the model for each individual can be done by rerunning the model once for each individual's row of data.

```
# change objective function
fsModel@objective <- mxFIMLObjective("theta", "mean",
names(expData), vector=TRUE)
# change data
fsModel@data <- mxData(exampleData[1,], "raw")
```

Defining and implementing the weight

The final step for this factor score model is to implement the expected posterior weighting. This can be done in OpenMx by including or wrapping our factor score model in a second model (essentially a "parent" or "container" model) whose function it is to multiply the objective function of the factor score model by the weighting algebra. When we run this full ML model, our factor score submodel will also be estimated. We can define the expected posterior weight as the normal density function at the location of the factor score estimate, which is shown in Equation 21.

$$f(\widehat{f_i}) = \frac{1}{(2\pi)^{k/2}|\sum|^{1/2}} \exp\left(-\frac{1}{2}(\widehat{f_i}-\mu)' \sum^{-1} (\widehat{f_i}-\mu)\right) \quad (21)$$

We can implement Equation 21 as an mxAlgebra that uses the matrices included in our factor score model. The following code creates our parent or container model, names it "Weighted Factor Score Model," adds our existing factor score model (the object fsModel), and defines the weight algebra ("weight").[5] That weighting algebra is then used to define the −2 log-likelihood in the algebra named "obj" and that algebra is specified as the objective function to

[5]This algebra could be simplified in our case, as the factor covariance structure is simply the number one. The full version is included for general purposes. Also note that the use of the existing factor mean matrix as our factor score estimator precludes the use of the factor mean μ in our normal density weight. To include nonzero factor means, the factor score and factor mean matrices must be made explicitly.

be minimized. Note that references to matrices and slots (e.g., objective) in the factor score model are prefaced with the name of that model and a period (e.g., Factor Score.).

```
# make the weighted factor score model
full Model <- mxModel("Weighted Factor Score Model",
fsModel,
mxAlgebra(1/((2*pi
)^(1/2) * sqrt(det(Factor Score.phi))) *
exp(-.5*(Factor Score.mu %&% solve(FactorScore.phi))),
name="weight"),
mxAlgebra(-2*log(weight %x% Factor Score.objective),
name="alg")
mxAlgebra Objective("alg")
)
# run the weighted factor score model
fsResults <- mxRunfull Model
```

Additional examples and a set of R functions for creating these models are available from the OpenMx website (http://openmx.psych.virginia.edu).

APPENDIX B

Missing at Random Data Data Generation

As described briefly in the method section and in Equations 13–15, missing at random data were generated by taking any condition's assigned probability of missingness δ and adjusting that probability up or down depending on any individual's observed data. This mechanism was implemented by creating a variable z_{ij}, which was a function of the first through $(j-1)^{th}$ elements of each individual's row of data y_i. This appendix describes the way that z_{ij} was created and its rationale.

The missingness propensity variable z_{ij} was calculated as the rescaled sum of a subset of the observed data y_i. For any column of data j where $j > 1$, z_{ij} was defined as the sum of any individual i's data from columns 1 to $j-1$. When $j = 1$, z_{ij} is equal to 0. This is given in Equation 22.

$$z_{ij} = \left\{ \begin{array}{ll} 0 & j=1 \\ \sum_{k=1}^{j-1} y_{ik}, & j>1 \end{array} \right\} \quad (22)$$

Missingness and missingness propensity (z_{ij}) were created sequentially across items, meaning that z_{ij} and missing values were created for all individuals on any data column j before any missingess treatment was applied to column $j + 1$. This sequential procedure is required to maintain that missing values depend only on nonmissing values of the data. If this was not done sequentially, data used to create any particular column of z_{ij} could be deleted as a missing value, making missingness depend on unobserved values and thus missing not at random.

The missingness propensity variable was then rescaled to a specific range, mean, and variance. This rescaling happened independently for each column. Each set of z_{ij} values for any value of j is rescaled to have a mean of 0 and standard deviation of.0167 ($\frac{1}{60}$), as shown in Equation 23. This scaling is

designed to place three standard deviations of z_{ij} within the range [−0.05, 0.05]; values more extreme than this were constrained as shown in Equation 24.

$$z_{ij}^{*}=\frac{z_{ij}-\bar{z}_{j}}{\sigma_{z_{j}} * 60} \quad (23)$$

$$P(R_{ij}|y_{ij}, MAR)=\left\{\begin{array}{lll} \delta+z_{ij}^{*}, & |z_{ij}^{*}| & < \ 0.05 \\ \delta+0.05, & z_{ij}^{*} & \geq \ 0.05 \\ \delta-0.05, & z_{ij}^{*} & \leq \ -0.05 \end{array}\right\} \quad (24)$$

Although the restriction of the scale and range of is admittedly arbitrary, it was selected to reflect the simulation design. Condition-wide missingness probability (δ) was selected from values of 0.05, 0.10, 0.15, 0.20, 0.25, and 0.30. Restricting the range of z_{ij} to [−0.05, 0.05] allowed the missingness scale to stay invariant over the different missingness conditions without exceeding the lower bound of 0 in the δ = 0:05 condition.

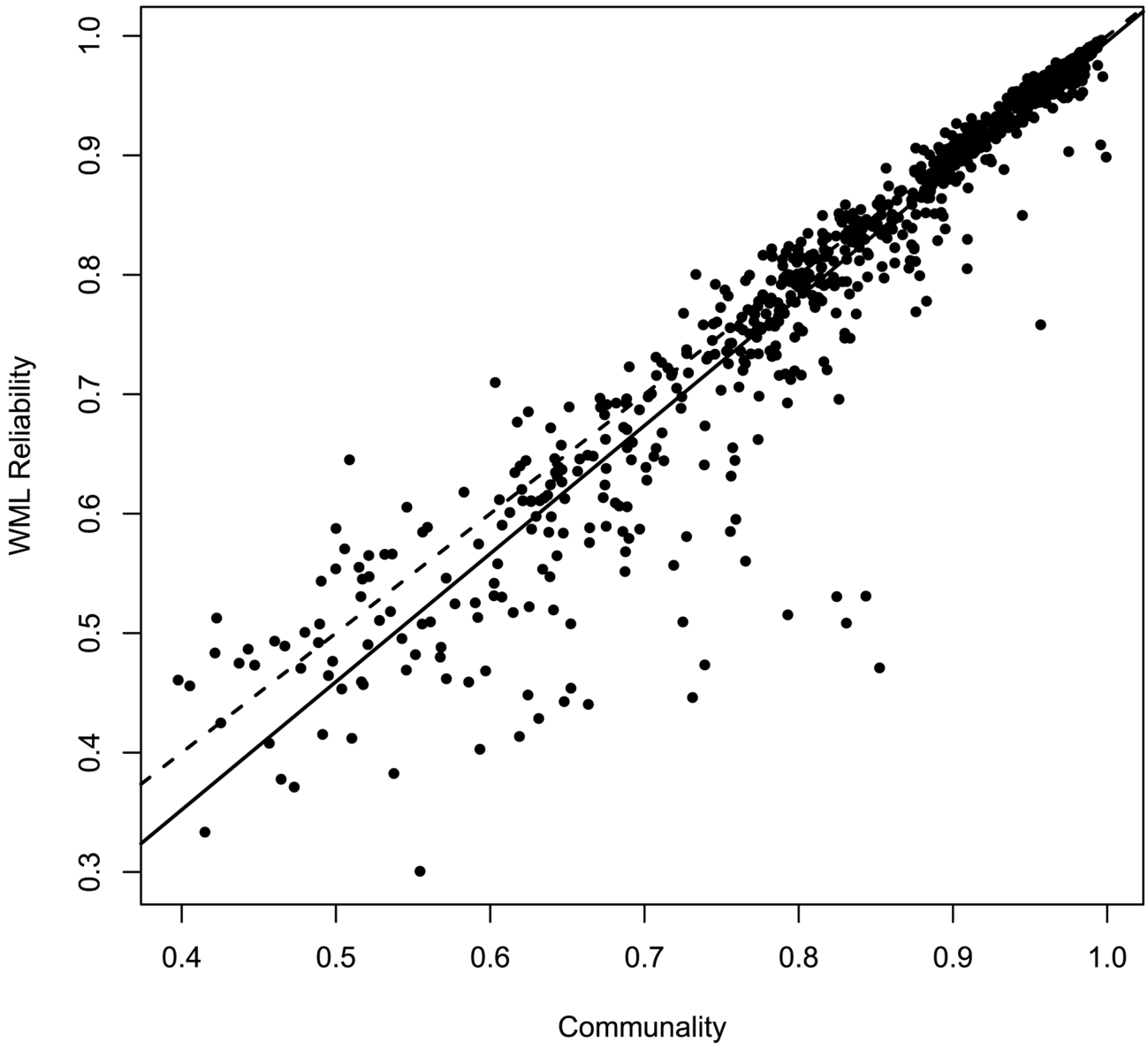

FIGURE 1.
Factor score-communality relationship under complete data. *Note.* Relationship between communality (x-axis) and factor score reliability (y-axis) for complete data.

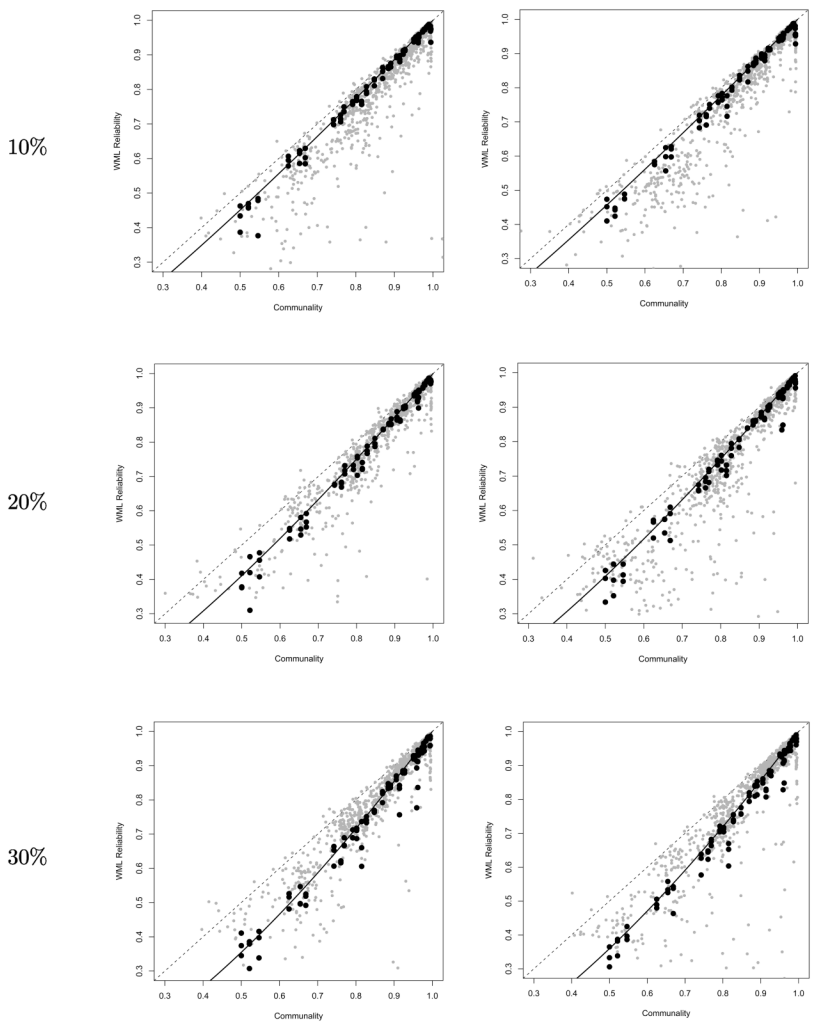

FIGURE 2.
Factor score reliability under missing data.

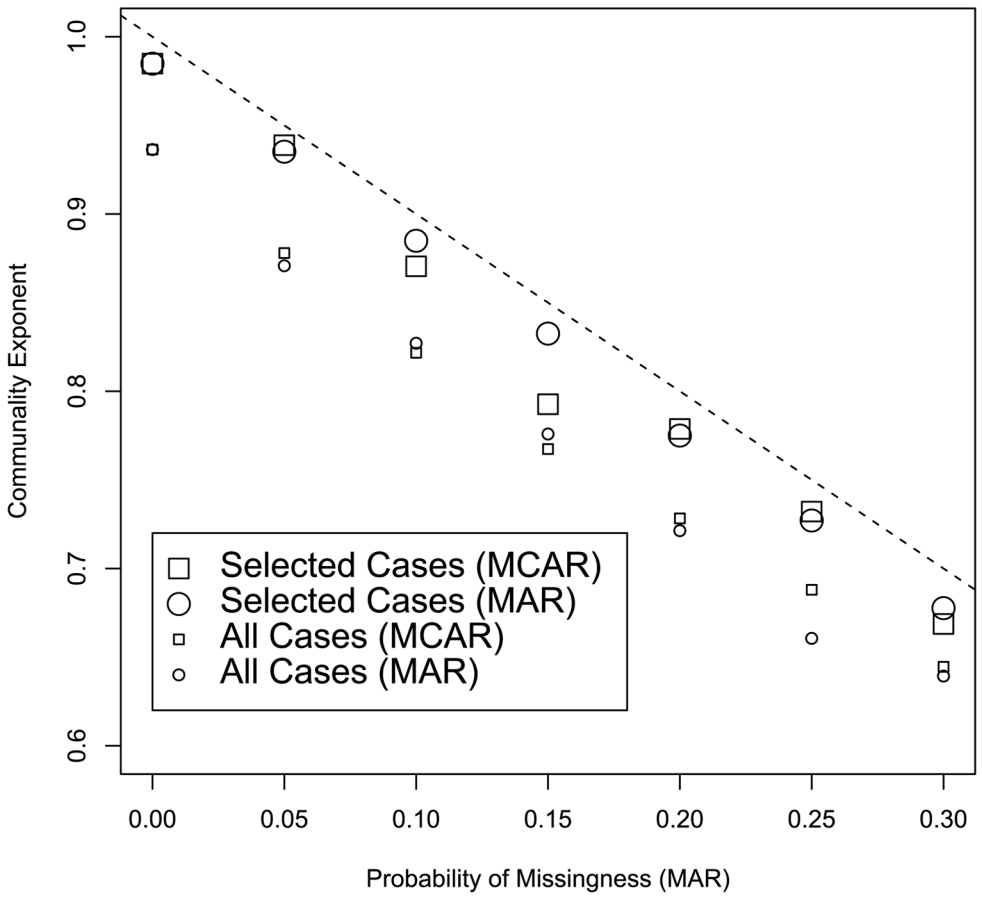

FIGURE 3.

Factor score-communality relationship under missing data. *Note*. Relationship between factor score reliability and communality under missingness. Exponent refers to ratio of *log* (*communality*) to *log*(*WML reliability*), which can also be expressed by the equation $communality = (WML\ reliability)^{\beta}$. Small dots represent all simulation conditions; larger dots represent conditions with greater than three variables and 100 individuals.

TABLE 1

Simulation Parameters

Parameter	Levels
Items	3, 5, 10, 20
Factor loadings (*M*)	0.5, 0.7, 0.9
Factor loadings (*SD*)	0.0, 0.2, 0.4
n	100, 200, 500
Iterations	10

TABLE 2

Mean Factor Score Reliability as a Function of Missing Data Mechanism and Number of Items

		Items			
Missingness	**Method**	**3**	**5**	**10**	**20**
Complete	Sum scores	0.717	0.793	0.878	0.932
	Bartlett method	0.731	0.811	0.893	0.941
	Unweighted ML	0.734	0.818	0.899	0.945
	Full ML	0.734	0.818	0.899	0.945
MCAR	Sum scores	0.666	0.752	0.852	0.917
	Bartlett method	0.658	0.754	0.853	0.919
	Unweighted ML	0.653	0.767	0.873	0.927
	Full ML	0.676	0.780	0.874	0.927
MAR	Sum scores	0.662	0.753	0.853	0.917
	Bartlett method	0.653	0.753	0.853	0.918
	Unweighted ML	0.648	0.768	0.874	0.926
	Full ML	0.673	0.780	0.876	0.926

Note. Mean reliability across complete data conditions for each of the four studied factor score methods given data structure and number of items. ML = maximum likelihood; MCAR = missing completely at random; MAR = missing at random.

Multivariate Behav Res. Author manuscript; available in PMC 2013 September 16.

TABLE 3

Predicting Reliability From Simulation Parameters

ML Factor Scores	Estimate	SE	p	R_p^2
Intercept	0.240	0.033	< .001	
n/100	0.011	0.039	0.788	0.000
Items	0.011	0.001	< .001	0.231
Mean loading	0.691	0.041	< .001	0.560
SD loading	0.089	0.041	< .001	0.009

Note. Regression of mean full maximum likelihood (ML) reliability for each complete data condition on simulation parameters. $R_p^2 = \text{partial } R^2$. Model R^2 =.801. Model R^2 when all cases (rather than cell means) used for regression =.773.

TABLE 4

Predicting Reliability From Communality

ML Factor Scores	Estimate	SE	p	R_p^2
Intercept	−0.052	0.006	< .001	
Communality	1.053	0.007	< .001	0.996

Note. Regression of mean full maximum likelihood (ML) reliability for each complete data condition on communality. $R_p^2 = \text{partial } R^2$. Model R^2 when all cases (rather than cell means) used for regression =.906.

TABLE 5

Predicting Nicotine Dependence From CHRNA Receptor Gene

	Complete Data			25% Missingness		
Method	**β**	χ^2_1	***p***	**β**	χ^2_1	***p***
Sum scores	.041	4.571	.033	.040	4.118	.042
Bartlett	.050	6.411	.011	.050	6.401	.011
Unweighted ML	.053	7.458	.006	.047	5.875	.015
Full ML	.053	7.458	.006	.051	6.871	.009
Full SEM model	.060	7.460	.006	.060	6.783	.009

Note. Standardized regressions of factor scores for a nicotine dependence factor on the rs16969968 SNP from the CHRNA receptor gene. Sum scores, Bartlett scores, and both types of ML estimates are estimated both with complete data and with 25% of responses deleted according to the MAR procedure described in the simulation. Full SEM model describes the size of the regression effect from a factor model fit to the item-level data. χ^2 and *p* values determined by likelihood ratio test.