

A Bayesian Approach for Estimating Mediation Effects with Missing Data

Craig K. Enders

Arizona State University

Amanda J. Fairchild

University of South Carolina

David P. MacKinnon

Arizona State University

Author Note

Craig K. Enders, Department of Psychology, Arizona State University. Amanda J. Fairchild, Department of Psychology, University of South Carolina. David P.

MacKinnon, Department of Psychology, Arizona State University.

Correspondence concerning this article should be addressed to Craig K. Enders, Department of Psychology, Arizona State University, Box 871104, Tempe, AZ, 85287-1104. Email: craig.enders@asu.edu.

This article was supported in part by National Institute on Drug Abuse grant DA09757 (David MacKinnon, PI).

Abstract

Methodologists have developed mediation analysis techniques for a broad range of substantive applications, yet methods for estimating mediating mechanisms with missing data have been understudied. This study outlined a general Bayesian missing data handling approach that can accommodate mediation analyses with any number of manifest variables. Computer simulation studies showed that the Bayesian approach produced frequentist coverage rates and power estimates that were comparable to those of maximum likelihood with the bias-corrected bootstrap. We share a SAS macro that implements Bayesian estimation and use two data analysis examples to demonstrate its use.

Keywords. Mediation, indirect effects, missing data, Bayesian estimation, bias corrected bootstrap, Sobel test.

A Bayesian Approach for Estimating Mediation Effects with Missing Data

The evaluation of mediating mechanisms has become a critical element of behavioral science research, not only to assess whether (and how) interventions achieve their effects, but also more broadly to understand the causes of behavioral change. The significance of mediation hypotheses is evident in a variety of substantive domains ranging from drug prevention (e.g., perceived social norms, a psychological mediator, transmits the impact of a prevention program on adolescent drug use; Liu, Flay, et al., 2009), to health promotion (e.g., knowledge of fruit and vegetable benefits, a behavioral mediator, transmits the effect of a health promotion program on weight loss; Elliot, Goldberg, Kuehl, Moe, Breger, & Pickering, 2007), to epidemiology (e.g., weight-related biomarkers, a biological mediator, transmit the effect of a diet and stress reduction intervention on tumor markers of prostate cancer; Saxe, Major, Westerberg, Khandrika, & Downs, 2008). The wide appeal of mediation analyses continues to drive the development of new analytic procedures for assessing mediation effects.

Methodologists have developed mediation analysis techniques for a broad range of substantive applications. However, methods for estimating mediating mechanisms with missing data have been understudied. The purpose of this study is to extend Yuan and MacKinnon's (2009) work on Bayesian mediation analyses to the missing data context. Specifically, we outline a Bayesian parameter simulation approach and provide a SAS macro for performing Bayesian mediation analyses with any number of manifest variables. Bayesian mediation analyses provide several advantages over conventional estimation methods. First, the Bayesian approach should yield more accurate inferences than frequentist significance testing approaches because it does not require distributional

assumptions. This advantage is particularly important because the sampling distribution of the mediated effect can be quite nonnormal. Second, the Bayesian paradigm offers a more interpretable interval estimate than the frequentist approach, providing a credible interval in which the probability of the population parameter falling within the bounds is specified. Third, it is possible to improve the power to detect mediation effects with missing data by incorporating prior information about the indirect effect. Finally, the parameter simulation approach that we demonstrate is straightforward to implement in SAS.

The organization of the manuscript is as follows. The paper begins with a review of both the single mediator model and Bayes' theorem. Next, we describe the use of a Markov chain Monte Carlo algorithm known as data augmentation to simulate the distribution of a mediation effect. Having outlined the procedural details, we use a data analysis example to demonstrate Bayesian estimation, and we use computer simulations to study its performance. Finally, we show how to perform a Bayesian analysis that incorporates prior knowledge about a mediation effect.

Mediation Analysis

A mediating variable (M) conveys the influence of an independent variable (X) on a dependent variable (Y), thereby illustrating the mechanism through which the two variables are related (Baron & Kenny, 1986; Judd & Kenny, 1981; MacKinnon, 2008). The basic mediation model posits that X influences M , which in turn influences Y , as follows

$$\hat{M} = \alpha X \quad (1)$$

$$\hat{Y} = \tau'X + \beta M \quad (2)$$

where α is the slope coefficient from the regression of M on X , τ' is the partial coefficient from the regression of Y on X controlling for M , and β is the slope from the regression of Y on M controlling for X . For simplicity, we omit the intercepts from the previous regression equations because these parameters do not contribute to the mediation effect.

Methodologists have proposed a variety of mediation tests (e.g., causal steps, Baron & Kenny, 1986; the difference in coefficients $\tau - \tau'$, MacKinnon, Warsi, & Dwyer, 1995), but contemporary work suggests that the product of coefficients is the most flexible estimator because it extends to complex scenarios that include multiple mediators or outcomes, multilevel data structures, and categorical outcomes (e.g., MacKinnon, Lockwood, Brown, Wang, & Hoffman, 2007). The logic of the product of the coefficients estimator derives from path analysis, where the total effect of X on Y is decomposed into indirect and direct components (Alwin & Hauser, 1975). This decomposition posits that the total effect of X on Y is comprised of a portion that is transmitted through the intervening variable (i.e., the indirect effect) and a portion that is not impacted by the intervening variable (i.e., the direct effect). This decomposition is given by

$$\tau = \tau' + \alpha\beta \quad (3)$$

where τ is the total effect of X on Y , τ' is the direct effect, and $\alpha\beta$ is the indirect effect. The indirect effect, or product of coefficients estimator, is the product of the regression slope that relates the independent variable to the mediator (i.e., the α path) and the partial regression coefficient that relates the mediator to the outcome (i.e., the β path). We henceforth refer to this estimator as $\alpha\beta$.

Researchers routinely use a normal-theory standard error derived from the multivariate delta method to evaluate the statistical significance of $\alpha\beta$ (e.g., Sobel, 1982). Dividing $\alpha\beta$ by a normal theory standard error yields a Wald z statistic, and referencing this test statistic to a standard normal distribution gives a probability value for the mediated effect. Recent work has shown, however, that normal-theory standard errors are inaccurate because the product of two normally distributed variables is not necessarily normally distributed; the sampling distribution of $\alpha\beta$ can be markedly asymmetric and kurtotic even when the α and β regression coefficients have normal sampling distributions (MacKinnon, Fritz, Williams, & Lockwood, 2007; MacKinnon, Lockwood, & Williams, 2004; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; Shrout & Bolger, 2002; Williams & MacKinnon, 2008). This distributional violation is especially problematic in small samples and with small effect sizes. The consequence of applying normal-theory significance tests is that the probability value from the Wald z test is conservative, limiting the power to detect mediation effects and increasing Type II error rates. The methodological literature currently favors asymmetric confidence limits based either on the theoretical sampling distribution of the product (MacKinnon et al., 2007; Meeker, Cornwell, & Aroian, 1981) or on empirical resampling methods (Efron & Tibshirani, 1993) because these approaches improve the accuracy of significance tests.

The Bayesian approach that we outline in this manuscript naturally yields asymmetric intervals as a byproduct of estimation.

Mediation Analyses with Missing Data

The last decade has seen a noticeable shift to missing data techniques that assume a missing at random (MAR) mechanism, such that the probability of missing data on an incomplete variable is unrelated to the would-be values of that variable (e.g., other variables in the data predict missingness). Maximum likelihood estimation and multiple imputation are currently the principal MAR-based methods in social science applications, and the Bayesian approach that we outline is a third option. Maximum likelihood estimation largely accommodates familiar complete-data mediation procedures. For example, multivariate delta standard errors (i.e., the Sobel test) and asymmetric confidence limits based on the bootstrap are available in software packages such as Mplus.

Multiple imputation is arguably less flexible for mediation analyses. Standard imputation procedures (e.g., draw m imputed data sets, perform a mediation analysis m times, pool the estimates and standard errors) encourage researchers to implement and pool delta method standard errors, and there is no reason to expect the performance of this approach to improve with incomplete data. Unfortunately, bootstrap procedures that methodologists currently favor do not translate well to multiple imputation. The most natural way to implement the bootstrap is to impute the data then apply the bootstrap to each imputed data set (i.e., impute-then-bootstrap). The procedural steps are as follows: (a) generate m imputed data sets, (b) draw b bootstrap samples from each data set, (c) form m empirical sampling distributions for $\alpha\beta$ by fitting the mediation model to the b

bootstrap samples from each imputed data set, and (d) find the $\alpha\beta$ values that correspond to the .025 and .975 quantiles of each empirical sampling distribution, $q_{.025}^m$ and $q_{.975}^m$, respectively. Because each of the m sets of bootstrap samples originates from a complete parent data set, averaging these quantiles or otherwise basing inferences on the bootstrap samples is inappropriate because the empirical sampling distributions reflect the variation of complete-data $\alpha\beta$ estimates (i.e., the bootstrap sampling distributions are narrower than their missing-data counterparts).

A second way to implement the bootstrap with imputation is to draw b samples from the incomplete data set and impute each bootstrap sample (i.e., bootstrap-then-impute). The procedural steps are as follows: (a) draw b bootstrap samples from the incomplete parent data set, (b) generate $m \geq 1$ imputed data set per sample, (c) estimate $\alpha\beta$ from each imputed data set, and (d) form an empirical sampling distribution of the $\alpha\beta$ estimates (or of the average of the m $\alpha\beta$ estimates if $m > 1$). Again, the .025 and .975 quantiles of this empirical sampling distribution define a 95% confidence interval that is the basis for statistical inferences. Unlike impute-then-bootstrap, this procedure should yield an appropriate empirical sampling distribution because an incomplete parent data set generates the bootstrap samples. However, bootstrap-then-impute is computationally intensive and difficult to implement in existing software packages.

The previous discussion highlights the fact that combining imputation and the bootstrap is, at best, a cumbersome solution for mediation analyses with missing data. We believe that a Bayesian framework provides a natural solution for mediation analyses because it yields a 95% confidence interval (technically, a 95% credible interval) based on the quantiles of an empirical distribution of estimates. This is precisely what

researchers are striving for with the bootstrap. Further, the simple SAS macro that we present later in the manuscript provides a convenient method for estimating a wide range of mediation models (e.g., single- or multiple-mediator models with any number of predictors or outcomes). Coupled with the ability to incorporate prior information (e.g., mediation estimates from pilot data or a published study), we believe that Bayesian mediation is an important tool for researchers.

Bayesian Estimation

This section briefly outlines Bayesian estimation. A number of resources provide accessible introductions to the Bayesian paradigm (Bolstad, 2007; Enders, 2010; Yuan & MacKinnon, 2009), and Bayesian texts provide additional technical details (Gelman, Carlin, Stern, & Rubin, 2009). The classical frequentist approach that predominates the social sciences defines a parameter as a fixed but unknown value in the population. Repeated sampling yields a distribution of estimates that vary around the true population value. In contrast, the Bayesian paradigm views the parameter itself as a random variable that has a distribution (i.e., there is no single true parameter value in the population). The basic goal of a Bayesian analysis is to use a priori information and the data to describe the shape of this parameter distribution (called a posterior). The prior information takes the form of a probability distribution for the parameter of interest, and a likelihood function quantifies the data's evidence about the parameter. Collectively, these two data sources define a posterior distribution that describes the relative probability of different parameter values.

Bayes' theorem provides the mathematical machinery behind Bayesian estimation. The theorem is

$$p(\theta|Y) = \frac{p(\theta)p(Y|\theta)}{p(Y)} \quad (4)$$

where $p(\theta|Y)$ is the posterior distribution of a parameter θ , Y is the data, $p(\theta)$ is the prior distribution of θ , $p(Y|\theta)$ is the likelihood function, and $p(Y)$ is the marginal distribution of the data. Because the denominator of theorem is an unnecessary scaling factor that makes the area under the posterior integrate (i.e., sum) to one, the key part of the theorem reduces to

$$p(\theta|Y) \propto p(\theta)p(Y|\theta). \quad (5)$$

In words, Equation 5 says that the relative probability of a particular parameter value is proportional to its prior distribution times the likelihood function. Conceptually, the posterior distribution is a weighted likelihood function, where the prior probabilities increase or decrease the height of the likelihood. For example, if the prior assigns a high probability to a particular value of θ , the corresponding point on the posterior is more elevated than the likelihood function. Conversely, if the prior specifies a low probability to a particular value of θ , the posterior distribution relies more heavily on the likelihood function.

There are at least two ways to apply the previous ideas to a mediation analysis. In the complete-data context, Yuan and MacKinnon (2009) applied Bayes' theorem to the α and β parameters that define of the product of coefficients estimator. Although this approach is also applicable to missing data analyses, we propose an equivalent procedure

that instead applies Bayesian principles to a covariance matrix. Working with a covariance matrix is convenient because it provides a straightforward mechanism for incorporating prior information about a mediation effect. As we show below, two quantities define the prior distribution of a covariance matrix: an a priori estimate of Σ and a degrees of freedom value. Because estimates of a covariance matrix (or alternatively, a correlation matrix) are often available from pilot data, a published study, or a meta-analysis, we believe that working with Σ is often easier than working directly with α and β . The remainder of this section describes the application of Bayesian estimation to a covariance matrix, and interested readers can consult Yuan and MacKinnon (2009) for details on applying Bayesian methods to regression coefficients.

The first step of a Bayesian analysis is to specify a priori probability distribution for the covariance matrix. Assuming normally distributed population data, this distribution belongs to the inverse Wishart family, a multivariate generalization of the chi-square distribution. More specifically, the prior distribution of a covariance matrix is

$$p(\Sigma) \sim W^{-1}(df_P, \Lambda_P^{-1}) \quad (6)$$

where W^{-1} denotes the inverse Wishart, df_P is the degrees of freedom, and Λ_P is a sum of squares and cross products matrix. In words, Equation 6 describes a researcher's a priori beliefs about the relative probability of different covariance matrix estimates arising from normally distributed population data. The parameters df_P and Λ_P are so-called hyperparameters that define the center (i.e., the location) and the spread (i.e., the scale) of the probability distribution, respectively.

As noted previously, a historical data source can determine the prior distribution's hyperparameters. For example, if a researcher had access to pilot data that employed identical measures of X , M , and Y , the sum of squares and cross products matrix from the pilot data can provide Λ_P . Similarly, the covariance matrix from a published study or a meta-analysis can provide the necessary matrix, as follows

$$\Lambda_P = (N - 1)\Sigma_P \quad (7)$$

provided that the published study used the same set of measures as your study. In the likely event that the prior data source employed measures with different metrics, using a correlation matrix to specify the a priori associations among the variables is an alternative solution. The following equation provides the necessary transformation

$$\Lambda_P = (N - 1)\Sigma_P = (N - 1)\hat{\mathbf{S}}\mathbf{R}_P\hat{\mathbf{S}} \quad (8)$$

where \mathbf{R}_P is the prior correlation matrix, and $\hat{\mathbf{S}}$ is a diagonal matrix that contains standard deviation estimates from the data (e.g., MAR-based maximum likelihood estimates). estimates of the standard deviations. Pre- and post-multiplying the correlation matrix by $\hat{\mathbf{S}}$ scales the correlations to the metric of the data.

As explained previously, a Bayesian analysis is essentially a weighted estimation problem where information from the prior and the data combine to produce an inference about the parameters of interest. The df_P value in Equation 6 quantifies the influence of the prior distribution on the sample size metric. For example, specifying a value of $df_P =$

20 is akin to saying that the prior distribution contributes 20 data points worth of information to the estimation process. If a researcher had a great deal of faith in the prior estimate of the covariance matrix, specifying a larger value of df_p would be defensible. Conversely, if a researcher had relatively little faith in the prior (e.g., because the historical data source used a slightly different population of participants, employed different measures, etc.), downweighting the prior with a small df_p value would be appropriate. In the absence of good a priori information about the covariance matrix, a researcher can specify a non-informative (i.e., diffuse) prior distribution that effectively contributes no information to estimation. In this situation, the data solely determine the final analysis results. Although it may appear to defeat the purpose of employing Bayesian methodology, specifying a non-informative prior is still quite useful because it provides a mechanism for evaluating mediation effects without invoking distributional assumptions for $\alpha\beta$. We demonstrate this point later in the manuscript.

In the Bayesian framework, the prior and the data combine to produce an updated distribution that describes the relative probability of different covariance matrix estimates. This posterior distribution is also an inverse Wishart

$$p(\Sigma|\mathbf{Y}) \sim W^{-1}(df, \Lambda^{-1}) \quad (9)$$

where a degrees of freedom and sum of squares and cross products matrix again define the center and the spread of the distribution, respectively. The degrees of freedom parameter is the sum of the degrees of freedom from the prior and the data, as follows

$$df = df_P + df_D = df_P + (N_D - 1) \quad (10)$$

and Λ similarly combines two sources of information¹.

$$\Lambda = \Lambda_P + \Lambda_D \quad (11)$$

Mediation Analyses via Data Augmentation

Deriving the posterior distribution in Equation 9 is difficult or impossible with missing data, and the observed-data distribution of $\alpha\beta$ is even more daunting. Markov chain Monte Carlo (MCMC) provides a simulation-based approach to estimating these distributions. Conceptually, an MCMC analysis is analogous to the bootstrap in the sense that it generates an empirical distribution for each parameter, but it does so by drawing a large number of parameter values from their respective posterior distributions; in the context of a missing data analysis, Schafer (1997) refers to this process as parameter simulation. The data augmentation algorithm (Schafer, 1997; Tanner & Wong, 1987) is the predominant MCMC algorithm for generating multiply imputed data sets, and the algorithm happens to be well suited for mediation analyses with incomplete data. As described below, data augmentation uses regression equations to impute the missing values, and it subsequently draws a new covariance matrix from its posterior distribution. In the context of a multiple imputation analysis, researchers are primarily interested in the imputed values and view the parameter draws as a superfluous byproduct of the MCMC

¹ When specifying a prior distribution for the variable means, Equation 10 has additional terms that depend on the prior means and the sample means (e.g., see Schafer, 1997, p. 152). For simplicity, we omit these terms because we believe that researchers will rarely have access to a prior data source that uses identical measures of X , M , and Y .

algorithm. In contrast, we are interested solely in the parameter draws, and the imputed data sets are the byproduct.

To begin, consider the use of data augmentation for simulating the posterior distribution of a covariance matrix. Data augmentation is a two-step algorithm that repeatedly cycles between an imputation step (I-step) and a posterior step (P-step). The I-step fills in the missing values with draws from a conditional distribution that depends on the observed data (i.e., the non-missing values) and the current estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. More formally, the I-step draws plausible score values from

$$\dot{Y}_t \sim p(Y_{\text{mis}} | Y_{\text{obs}}, \dot{\boldsymbol{\theta}}_{t-1}) \sim N(\hat{Y}_i, \sigma_\varepsilon^2) \quad (12)$$

where \dot{Y}_t represents an imputed value from cycle t (throughout the manuscript, we use a dot to denote a draw), Y_{mis} is the missing value, Y_{obs} represents the observed data, and $\dot{\boldsymbol{\theta}}_{t-1}$ contains the parameter draws (i.e., $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$) from the previous computational cycle. Procedurally, data augmentation uses multiple regression to simulate random draws from this conditional distribution. Specifically, the elements in $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ generate regression equations that predict the incomplete variables from the complete variables, such that each missing value is replaced by a draw from a normal distribution with a mean equal to the predicted score for case i and a variance equal to the residual variance from the regression of Y on the complete variables.

Having filled in the data, the P-step draws a new covariance matrix and mean vector from their respective complete-data posterior distributions. To begin, data augmentation uses the filled-in data set from the preceding I-step to estimate the variable

means and the sum of squares and cross products matrix, $\hat{\boldsymbol{\mu}}_t$ and $\hat{\boldsymbol{\Lambda}}_t$, respectively. Combined with the hyperparameters of the prior distribution (i.e., df_P and $\boldsymbol{\Lambda}_P$) these quantities define the posterior distribution of $\boldsymbol{\Sigma}$. To generate a new estimate of the covariance matrix, the algorithm uses Monte Carlo simulation to draw a k by k matrix of random numbers from the posterior distribution in Equation 9, where $\hat{\boldsymbol{\Lambda}}_t$ replaces $\boldsymbol{\Lambda}_D$. The Monte Carlo procedure yields a simulation-based estimate of $\boldsymbol{\Sigma}$ that randomly differs from the covariance matrix that produced the imputation regression coefficients at the preceding I-step. We denote this draw as $\dot{\boldsymbol{\Sigma}}_t$. The algorithm similarly generates a new set of means by drawing a k by 1 vector of random numbers from a multivariate normal distribution, the center and spread of which are defined by $\hat{\boldsymbol{\mu}}_t$ and $\dot{\boldsymbol{\Sigma}}_t$, respectively. We denote this draw as $\dot{\boldsymbol{\mu}}_t$. Having completed a single computational cycle, the data augmentation passes the simulated parameters to the next I-step, where the elements in $\dot{\boldsymbol{\mu}}_t$ and $\dot{\boldsymbol{\Sigma}}_t$ define the imputation regression equations for cycle $t + 1$.

The data augmentation algorithm repeatedly cycles between the I- and P-steps, often for thousands of iterations. The collection of $\dot{\boldsymbol{\mu}}_t$ and $\dot{\boldsymbol{\Sigma}}_t$ values forms an empirical posterior distribution for each element in the mean vector and the covariance matrix. Although the covariance matrix is not the substantive focus of a mediation analysis, the elements in $\dot{\boldsymbol{\Sigma}}$ define the mediation model parameters. For example, the components of the product of coefficients estimator for a single-mediator model are as follows

$$\hat{\alpha} = \hat{\sigma}_{XM} / \hat{\sigma}_X^2 \quad (13)$$

$$\hat{\beta} = (\hat{\sigma}_{MY}\hat{\sigma}_X^2 - \hat{\sigma}_{XM}\hat{\sigma}_{XY}) / (\hat{\sigma}_X^2\hat{\sigma}_M^2 - \hat{\sigma}_{XM}^2) \quad (14)$$

and the direct effect is given by

$$\hat{\tau}' = \hat{\sigma}_{MY}\hat{\sigma}_X^2 - \hat{\sigma}_{XM}\hat{\sigma}_{XY}/\hat{\sigma}_X^2\hat{\sigma}_M^2 - \hat{\sigma}_{XM}^2. \quad (15)$$

Although they may or may not be of interest, the covariance matrix elements also define the residual variances from the mediation model.

$$\hat{\sigma}_{\varepsilon_M}^2 = \hat{\sigma}_M^2 - \hat{\sigma}_{XM}^2/\hat{\sigma}_X^2 \quad (16)$$

$$\hat{\sigma}_{\varepsilon_Y}^2 = (\hat{\sigma}_X^2\hat{\sigma}_M^2\hat{\sigma}_Y^2 - \hat{\sigma}_X^2\hat{\sigma}_{MY} - \hat{\sigma}_M^2\hat{\sigma}_{XY} - \hat{\sigma}_Y^2\hat{\sigma}_{XM} + 2\hat{\sigma}_{XM}\hat{\sigma}_{XY}\hat{\sigma}_{MY})/(\hat{\sigma}_X^2\hat{\sigma}_M^2 - \hat{\sigma}_{XM}^2) \quad (17)$$

Analogous matrix expressions are applicable to multiple-mediator models.

Because each P-step yields values for the terms on the right side of the previous equations, data augmentation provides a simple mechanism for simulating draws from the posterior distribution of $\alpha\beta$. The steps are as follows: (a) generate a data augmentation chain comprised of several thousand (e.g., 10,000) computational cycles, (b) save the values in $\hat{\Sigma}_t$ at each P-step, and (c) use the elements in $\hat{\Sigma}_t$ to compute $\hat{\alpha}\hat{\beta}_t$ from each cycle. The collection of $\hat{\alpha}\hat{\beta}$ values forms a simulation-based posterior distribution for the mediation effect. Importantly, parameter simulation does not use the filled-in data sets to produce an inference about $\alpha\beta$; the imputations are a superfluous byproduct of estimation. Rather, the goal is to use familiar summary statistics to describe the center and spread of the posterior. For example, the median $\hat{\alpha}\hat{\beta}$ value describes the center of the

posterior distribution and effectively serves the same purpose as a frequentist point estimate. Although the posterior mean is sometimes a suitable point estimate, the posterior distribution of $\alpha\beta$ can be markedly nonnormal, making the median a better choice². Similarly, the posterior standard deviation is comparable to a frequentist standard error.

Finally, the $\alpha\beta$ values that correspond to the .025 and .975 quantiles of the posterior distribution, $q_{0.025}$ and $q_{0.975}$, define a credible interval that is analogous to a 95% confidence interval in the frequentist paradigm. The credible interval provides a straightforward mechanism for hypothesis testing. Consistent with standard practice, the mediation effect is statistically significant if this interval does not contain zero. Although the 95% credible interval functions like a frequentist confidence interval, it is important to note that its interpretation does not rely on a hypothetical process of drawing repeated samples from the population. Rather, the credible interval defines the range in which 95% of the parameter values fall. Importantly, because the 95% credible interval is based on the empirical quantiles of the posterior distribution, inferences do not rely on the (often) untenable assumption that the $\alpha\beta$ estimates follow a normal distribution. This is an important advantage over maximum likelihood and multiple imputation procedures that utilize multivariate delta method standard errors (MacKinnon et al., 2007; MacKinnon, Lockwood, & Williams, 2004; MacKinnon et al., 2002).

Data Analysis Example 1

² We compared the posterior mean and median from the computer simulations that we report in a subsequent section and found that the posterior median had a lower mean squared error when compared to the data-generating value of $\alpha\beta$.

To illustrate the parameter simulation procedure, we analyzed data from a study that sought to reduce the risk for cardiovascular disease in firefighters (Moe et al., 2002; Ranby, MacKinnon, Fairchild, Elliot, Kuehl, & Goldberg, 2011). Due to the extreme physical demands of their job, firefighters have a markedly elevated risk for cardiovascular disease. Accordingly, the PHLAME (“Promoting healthy lifestyles: Alternative models’ effects”) intervention aimed to decrease risk factors in this group by promoting healthy eating and exercise habits. For this example, we consider the impact of a team-based treatment program (X) on intent to eat fruit and vegetables (Y) through two putative mediators: the extent to which coworkers eat fruit and vegetables (M_1) and knowledge of healthy eating benefits (M_2). Additionally, we used two auxiliary variables in the imputation model, friends’ and partner’s consumption of fruit and vegetables and partner’s consumption of fruit and vegetables. Figure 1 shows a path diagram of the mediation model (to avoid clutter, the residual covariance between the mediators is omitted). Table 1 gives maximum likelihood descriptive statistics for the analysis variables.

Although multiple imputation programs typically output the parameter draws required for computing mediation model parameters, the computations require considerable data manipulation, particularly for models with multiple mediators or multiple outcomes. To simplify the process, we created a SAS macro program that fully automates the Bayesian analysis. The macro uses the MI procedure in SAS to implement parameter simulation and subsequently uses the IML procedure to compute the mediation model parameters. The full macro program is available at www.appliedmissingdata.com.

To illustrate the macro, Figure 2 shows the inputs for the data analysis example, with bold typeface denoting user-supplied values. The macro requires four sets of inputs. The first set of inputs consists of the file path for the raw data, the input variable list, and a numeric missing value code. We assume that the input text file is in free format and with a single placeholder value for missing data. Note that the input data may contain variables that are not used in the analysis. The second set of inputs contains information about the prior distribution. Leaving these entries blank, as they are in the figure, invokes a standard non-informative prior. Later in the manuscript we demonstrate how to specify an informative prior distribution. The third set of inputs specifies the analysis variables. Variables in the mediation model are specified as X , M , or Y variables. Although our example is a relatively simple two-mediator model, the macro is flexible and can accommodate single- or multiple-mediator models within any number of X , M , and Y variables. Auxiliary variables (e.g., correlates of incomplete variables or predictors of missingness) contribute to the imputation process but are not part of the mediation model. Specifying which variables are complete or missing is unnecessary, but it is important to note that the linear imputation model in PROC MI assumes that the incomplete variables are multivariate normal. At a minimum, this assumption precludes the use of incomplete categorical variables with more than two categories; if complete, such variables can be used as X variables or as auxiliary variables, but they must be dummy or effect coded. The final set of inputs specifies the burn-in period, the number of iterations, and a random number seed (the seed can be left blank, but specifying a value ensures that the program will return the same results each time it executes). The specifications in our

example instruct the macro to discard the initial set of 500 burn-in iterations and form the posterior distributions from the ensuing 10,000 iterations.

To facilitate convergence assessments, the macro produces trace plots of the mediation parameters from the burn-in period. A trace plot is a line graph that displays the iterations on the horizontal axis and the simulated parameter values on the vertical axis. To illustrate, Figure 3 shows a plot of the $\alpha\beta$ values for the indirect effect of the intervention on intentions via coworker eating habits. The absence of long-term upward or downward trends suggests that a stable posterior distribution generated the parameter values. For interested readers, a number of resources further address the use of trace plots as a convergence diagnostic (Enders, 2010; Schafer, 1997; Schafer & Olsen, 1998).

Turning to the posterior distribution, the macro produces a histogram and a kernel density plot for each mediated effect. To illustrate, Figure 4 shows a plot of the $\alpha\beta$ values for the indirect effect of the intervention on intentions via coworker eating habits. In our example, this distribution was based on draws from the 10,000 iterations following the burn-in period. As seen in the figure, the posterior distribution was nonnormal, with skewness of $S = .46$ and excess kurtosis of $K = .59$. Again, the shape of the posterior underscores the problem with applying significance tests that assume a normal distribution (and thus a symmetric confidence interval) for $\alpha\beta$ (e.g., tests based on the multivariate delta method).

In a Bayesian analysis, researchers can use straightforward summary statistics to describe and evaluate mediation effects. For example, the mean and the median describe the center of the posterior distribution in a manner that is analogous to a frequentist point estimate. Although both values are legitimate estimates of the mediation effect, the

median is generally a better estimate, particularly when the posterior distribution is nonnormal (e.g., when the sample size or the effect size is small). In lieu of a formal significance test, the parameter values that correspond to the .025 and .975 quantiles of the posterior distribution form a credible interval that is analogous to a 95% confidence interval in the frequentist paradigm. The macro produces a table that includes the posterior median and the 95% credible interval bounds for each mediated effect.

Figure 5 shows the macro output from the example analysis. Considering the indirect effect of the intervention on intentions via coworker eating habits, the posterior median was $Mdn = .083$, and the 95% credible interval limits were $q_{0.025} = .0010$ and $q_{0.975} = .193$. Because this interval does not contain zero, we can conclude that the mediation effect is statistically significant. The corresponding absolute proportion mediated effect size was .264, indicating that approximately 26% of the effect of the team intervention on intent to eat fruit and vegetables was mediated through the social norm of coworkers eating fruit and vegetables. Turning to the indirect effect of the intervention on intentions via knowledge of healthy eating benefits, the posterior median was $Mdn = .063$, and the 95% credible interval limits were $q_{0.025} = -.00007$ and $q_{0.975} = .162$, and the corresponding absolute proportion mediated was .203. From a frequentist perspective, this indirect effect is non-significant because the credible interval contained zero. Importantly, both credible intervals are asymmetric around the posterior median, which owes to the positive skew of the posterior distributions.

The interpretation of component path coefficients in the mediation model may be of particular interest in program evaluation work where one seeks to explore action and/or conceptual theory underlying a program (see Chen, 1990). The same summary

statistics detailed above can describe the posterior distributions of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\tau}'$ individually. As seen in Figure 5, the tabular output also includes numeric summaries for the component paths as well as for the total effect of X on Y .

Computer Simulations

We performed a series of Monte Carlo simulations to study the performance of Bayesian mediation. With one exception, the data-generating model was consistent with full mediation and did not include a direct effect of X on Y . We generated 5000 artificial data sets within each cell of a fully crossed design that manipulated three experimental conditions: the missing data mechanism (missing completely at random and missing at random), the sample size ($N = 50, 100, 300, \text{ and } 1000$), and the magnitude of the mediation effect. The mediation factor was comprised of five combinations of α and β values: $\alpha = 0$ and $\beta = 0$, $\alpha = .10$ and $\beta = .10$, $\alpha = .30$ and $\beta = .30$, $\alpha = .50$ and $\beta = .50$, and $\alpha = .30$ and $\beta = .10$. Because the simulation variables were standard normal, the α and β values are identical to Pearson correlations. Consequently, the mediation model parameters correspond to Cohen's (1988) effect size benchmarks (i.e., zero/zero, small/small, medium/medium, large/large, medium/small). Finally, note that the $\alpha = 0$ and $\beta = 0$ condition had a direct effect that was moderate in magnitude (i.e., $\tau' = .30$). The direct effect was zero in all other conditions.

We used the SAS IML procedure to generate three standard normal variables and subsequently used Cholesky decomposition to impose the desired correlation structure. Next, we imposed a 20% missing data rate on both M and Y . In the missing completely at random (MCAR) simulation, we used uniform random numbers to independently impose missing data, such that cases with $u_i < .20$ had a missing value on M or Y . In the MAR

simulation, the value of X determined missingness on M and Y . Beginning with the highest X value and working in descending order, we independently deleted M and Y scores with a .75 probability until 20% of the values on each variable were missing. This produced a situation where cases with high scores on X had higher missing data rates on M and Y . Note that we chose not to manipulate the amount of missing data because this design characteristic tends to produce predictable and uninteresting findings (e.g., as the missing data rate increases, power decreases). Rather, we chose a constant 20% missing data rate because we felt that it was extreme enough to expose any problems.

We used the MI procedure in SAS version 9.2 to implement parameter simulation. Using the standard non-informative prior distributions for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, we generated 10,500 cycles of data augmentation and saved the simulated parameter estimates from each P-step. Next, we used the formulas from Equations 13 through 15 to convert the elements in $\hat{\boldsymbol{\Sigma}}_t$ into mediation model parameters. For each of the 200,000 replications (2 missing data mechanisms by 4 sample size conditions by 5 effect size conditions by 5000 iterations), we discarded the parameter draws from the initial 500 burn-in cycles and formed an empirical distribution of 10,000 parameter draws.

To assess the relative performance of the Bayesian approach, we also used maximum likelihood missing data handling and multiple imputation to estimate the mediation effect. These are the principle MAR-based analysis approaches that enjoy widespread use in substantive applications. To implement maximum likelihood estimation, we input the raw data files to Mplus 6.1 and used the MODEL INDIRECT command to estimate $\alpha\beta$. Mplus offers a variety of significance testing options for mediation effects, and we examined three different possibilities: normal-theory maximum

likelihood with multivariate delta method (i.e., Sobel) standard errors, the bias corrected bootstrap, and the percentile bootstrap. Although the methodological literature has discounted significance tests based on normal-theory standard errors, we chose this option because it likely reflects the analytic practice of many substantive researchers, especially those who use statistical software packages such as SPSS or SAS.

We used the MI procedure in SAS version 9.2 to implement multiple imputation. Using the standard non-informative prior distributions for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, we generated 100 imputed data sets from a data augmentation chain with 200 burn-in and 200 between-imputation cycles; we chose the relatively large number of imputed data sets in order to maximize power of the subsequent significance tests (Graham, Olchowski, & Gilreath, 2007). We then used the REG procedure to estimate the regression coefficients from Equations 13 through 15 and subsequently computed $\alpha\beta$ and its normal-theory (i.e., Sobel) standard error from each of the 100 sets of imputations. Finally, we used Rubin's (1987) pooling rules to generate a point estimate, standard error, and confidence interval for each of the 200,000 artificial data sets.

The outcome variables for the simulations were relative bias, power, and confidence interval coverage. Because the Bayesian framework defines a parameter as a random variable rather than a fixed quantity, frequentist definitions of power and confidence interval coverage do not apply. Nevertheless, we felt that it was important to evaluate the Bayesian procedure from a frequentist perspective in order to make comparisons with other MAR-based procedures. To quantify power for the Bayesian analyses, we obtained the .025 and the .975 quantiles from each posterior distribution and defined power as the proportion of replications where the 95% credible interval did not

include zero. We applied the same procedure to the maximum likelihood bootstrap procedures. For maximum likelihood and multiple imputation with normal-theory test statistics, the proportion of replications within each design cell that produced a statistically significant $\alpha\beta$ coefficient served as an empirical estimate of power.

For Bayesian (and bootstrap) confidence interval coverage, we used the .025 and the .975 quantiles from each posterior (sampling) distribution to compute the proportion of the credible intervals within each design cell that included the population parameter from the data-generating model. For maximum likelihood and multiple imputation with normal-theory standard errors, coverage was the proportion of replications where the 95% confidence interval included the true population value. In the frequentist paradigm, confidence interval coverage value directly relates to the Type I error rate, such that 95% coverage corresponds to the nominal 5% Type I error rate, a 90% coverage rate corresponds to a twofold increase in Type I errors, and so on.

Simulation Results

As would be expected with MCAR and MAR mechanisms, all missing data handling approaches produced relatively accurate point estimates with no appreciable bias. Consequently, we limit our subsequent presentation to power and confidence interval coverage. Further, because the results from the MCAR and MAR simulations were virtually identical, we limit the presentation to the MAR condition.

Table 2 displays the coverage values for each design cell in the MAR simulation. The table illustrates that the effect size and the sample size influenced coverage rates and that the sample size moderated the impact of effect size, such that larger sample sizes yielded more accurate coverage. The $\alpha = 0$ and $\beta = 0$ condition produced coverage rates

that exceeded the nominal .95 rate at every sample size, indicating conservative inferences. The same was true for the condition where both coefficients had a small effect size (i.e., $\alpha = .10$ and $\beta = .10$), although the coverage rates approached the nominal level as the sample size increased. Finally, when either α or β (or both) had at least a medium effect size, Bayesian estimation and maximum likelihood estimation with bootstrap standard errors produced accurate coverage rates, particularly with sample sizes of 100 or larger. In contrast, coverage rates based on multivariate delta (i.e., Sobel) standard errors were not as accurate at smaller sample sizes and only approached the nominal level at larger N s. In the complete-data context, Yuan and MacKinnon (2009) compared Bayesian and frequentist coverage values and found a very similar pattern of results.

Table 3 gives empirical power estimates for the five methods. For Bayesian estimation and maximum likelihood estimation with bootstrap standard errors, these estimates reflect the proportion of replications where the credible (or confidence) interval did not contain zero, whereas the power values for the normal-theory test statistics reflect the proportion of replications that produced a statistically significant test statistic. As seen in the table, the bias corrected bootstrap produced the highest power values, followed by Bayesian estimation and the percentile bootstrap, respectively. Power values for the normal-theory (i.e., Sobel) test were noticeably lower, except in situations where the sample size or the effect size was large.

Collectively, the results in Tables 2 and 3 suggest that Bayesian parameter simulation yields coverage and power values that are comparable to the best maximum likelihood approach (maximum likelihood estimation with the bias corrected bootstrap).

As expected, Bayesian estimation was superior to normal-theory significance tests based on multivariate delta method standard errors. It is important to reiterate that we implemented standard non-informative prior distributions in the simulations. Bayesian estimation can produce a substantial power advantage (i.e., narrower confidence intervals) when using informative prior distributions (Yuan & MacKinnon, 2009). We address this topic in the next section.

Data Analysis Example 2

Thus far, we have focused on non-informative prior distributions. In the context of a complete-data mediation analysis, Yuan and MacKinnon (2009) showed that an informative prior could provide a substantial increase in statistical power. This is particularly useful for behavioral science applications where mediation effects are often modest in magnitude. In this section, we use our SAS macro to illustrate a mediation analysis with an informative prior distribution. To keep the example simple, we consider a single-mediator model involving the indirect effect of the team intervention on intentions via coworker eating habits. Further, we omit the auxiliary variables from the previous example.

As described previously, implementing an informative prior distribution requires an a priori estimate of the covariance or correlation matrix and a degrees of freedom value (e.g., from pilot data, a published study, or a meta-analysis). For the purposes of illustration, suppose pilot data produced the following covariance matrix and mean vector.

$$\hat{\Sigma} = \begin{bmatrix} .250 & .188 & .075 \\ .188 & 2.250 & .450 \\ .075 & .450 & 2.250 \end{bmatrix}$$

$$\hat{\mu} = \begin{bmatrix} .50 \\ 3.50 \\ 3.25 \end{bmatrix}$$

The MI procedure (and thus our macro) uses a `_TYPE_ = COV` data set containing estimates $\hat{\Sigma}$ and $\hat{\mu}$ to define the prior distributions (the program automatically converts the matrices into the necessary hyperparameters). Our macro program requires a file path for this data set (in text format) as well as a corresponding variable list. To illustrate, Figure 6 shows the macro inputs for the analysis example, and Figure 7 shows the contents of the `_TYPE_ = COV` text data set containing the prior parameters. Importantly, the variable list for the prior data must have `_TYPE_` and `_NAME_` as the first two variables, and the format of the data file itself should not deviate from that in Figure 7. The `N` and `N_MEAN` rows of the input data set contain the degrees of freedom values for the covariance matrix and mean vector, respectively. Although the decision is subjective, we arbitrarily assigned 30 degrees of freedom to the prior distribution, which effectively meant that the prior contributed roughly 30 as much information as the data. Finally, note that the values in the `MEAN` and `N_MEAN` rows can be set at zero if mean estimates are not available. Such would be the case when using Equation 8 to convert a correlation matrix from a published study that uses measures with different metrics.

Consistent with the earlier example, we generated a 10,500-cycle data augmentation chain and discarded the initial set of 500 burn-in iterations. For comparison purposes, we estimated the single-mediator model with a standard non-

informative prior. The posterior median was $Mdn = .079$, and the 95% credible interval was bounded by values of $q_{0.025} = -.003$ and $q_{0.975} = .192$. Again, notice that this interval was asymmetric around center of the distribution because the posterior was positively skewed. Importantly, implementing an informative prior reduced the variance of the posterior distribution and therefore narrowed width of the credible interval (i.e., increased power). Specifically, the posterior median was $Mdn = .085$, and the 95% credible interval was bounded by values of $q_{0.025} = .011$ and $q_{0.975} = .188$. Notice that the mediation effect was statistically different from zero because credible interval no longer contained zero. To visually illustrate the analysis results, Figure 8 shows kernel density plots from the single-mediator model. The plots clearly demonstrate that the posterior distribution based on an informative prior is narrower than that of the non-informative prior. In frequentist terms, the reduction in variance translates into greater power.

Discussion

Assessing mediating mechanisms is a critical component of behavioral science research. Methodologists have developed mediation analysis techniques for a broad range of substantive applications, but methods for missing data have thus far gone unstudied. Consequently, the purpose of this study was to outline a Bayesian approach for estimating and evaluating mediation effects with missing data. Although our approach is closely related to that of Yuan and MacKinnon (2009) in the complete-data context, but it applies Bayesian machinery to a covariance matrix rather than to the α and β coefficients themselves. In our view, specifying the mediation model in terms of Σ has two important advantages. First, we can use normal-theory multiple imputation software

(e.g., our macro uses PROC MI in SAS) to implement the procedure. Programs such as PROC MI output the posterior draws that are a byproduct of imputation, and converting the elements of $\hat{\Sigma}$ to mediation model parameters is straightforward (e.g., using Equations 13 through 15 or matrix versions of those expressions). Second, and most importantly, parameterizing the mediation model as a function of a saturated covariance matrix simplifies the process of specifying a prior distribution because researchers need only have an a priori estimate of Σ or R . Suitable estimates are often available from pilot data, published research studies, or meta-analyses. Having a convenient mechanism for specifying prior distributions is particularly important because it can partially compensate for the reduced precision that is inherent with missing data.

Even with a non-informative prior, our simulation results suggest that Bayesian estimation yields frequentist coverage values and power estimates that are comparable to maximum likelihood estimation with the bias corrected bootstrap; based on the complete-data literature, we expected the bias corrected bootstrap to be the gold standard against which to compare the Bayesian method. As expected, Bayesian estimation was superior to normal-theory significance tests, particularly when the sample size or the effect size of either the α or β paths was small. These findings align with those of Yuan and MacKinnon (2009) in the complete-data context. Although we did not pursue simulations with informative prior distributions, past research and statistical theory predict that Bayesian estimation would produce narrower confidence intervals and greater precision than frequentist alternatives, including the bootstrap. Our second analysis example supports this assertion.

Our study suggests a number of avenues for future research. First, the lack of existing research on this topic prompted us to limit the simulations to a single-mediator model. Although this model is exceedingly common in the behavioral sciences, future studies should investigate models with multiple mediators and outcomes. Our SAS macro is general and can accommodate these additional complexities. Second, our approach assumes that the incomplete variables are multivariate normal and it further assumes that the M and Y variables are continuous, regardless of whether they have missing data. Flexible MCMC methods are becoming available for mixtures of categorical and continuous variables (Enders, 2012; Goldstein, Carpenter, Kenward, & Levin, 2009; van Buuren, 2007), and these could easily be adapted to accommodate mediation models. This would be a fruitful avenue for future research, particularly given the recent interest in mediation models with categorical outcomes (MacKinnon et al., 2007). Fourth, recent research suggests that the bias corrected bootstrap yields elevated Type I errors under particular sample size and effect size configurations (Fritz, Taylor, & MacKinnon, 2012). Although we did not observe between-method differences in Type I errors (for brevity, we did not report the Type I error rates from the $\alpha = 0$ and $\beta = 0$ design cells), we did not examine the combination of conditions that produce elevated Type I errors. Future studies should explore the differences between Bayesian estimation and the bias corrected bootstrap in these problematic scenarios. Finally, all computer simulations are necessarily limited in scope, and ours was no exception. Although we examined a wide range of sample size and effect size conditions (the primary factors that influence the distribution of the product), future studies could extend our work by

investigating different distribution shapes, missing data rates, and missing data mechanisms.

In sum, Bayesian estimation provides a straightforward mechanism for estimating and evaluating mediation effects with missing data. Our simulations suggest that the Bayesian approach should yield comparable results to the best possible maximum likelihood analysis (maximum likelihood with bias corrected bootstrap). When researchers have access to a priori information about a mediation effect (e.g., from pilot data, a published study, or a meta-analysis), Bayesian estimation can provide a substantial increase in power. We believe that specifying the prior information in terms of a covariance or a correlation matrix will allow researchers to take advantage of this important benefit.

References

- Alwin, D.F. & Hauser, R.M. (1975). The decomposition of effects in path analysis. *American Sociological Review*, 40, 37-47.
- Baron, R.M. & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- Bolstad, W.M. (2007). *Introduction to Bayesian statistics* (2nd Ed.). New York: Wiley.
- Chen, H.T. (1990). *Theory-driven evaluations*. Newbury Park, CA: Sage.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Erlbaum.
- Efron, B., & Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall, New York.
- Elliot, D.L., Goldberg, L., Kuehl, K.S., Moe, E.L., Breger, R.K.R., Pickering, M.A. (2007). The PHLAME (Promoting Healthy Lifestyles: Alternative Models' Effects) firefighter study: Outcomes of two models of behavior change. *Journal of Occupational and Environmental Medicine*, 49(2), 204-213.
- Enders, C.K. (2012). *A chained equations approach for imputing single and multilevel data with categorical and continuous variables*. Manuscript in preparation.
- Enders, C.K. (2010). *Applied missing data analysis*. New York: Guilford Press.
- Fritz, M.S., Taylor, A.B., & MacKinnon, D.P. (2012). Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research*, 47, 61-87.

- Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (2009). *Bayesian data analysis* (2nd Ed.) Boca Raton, FL: Chapman & Hall.
- Goldstein, H., Carpenter, J., Kenward, M.G., & Levin, K.A. (2009). Multilevel models with multivariate mixed response types. *Statistical Modelling*, 9, 173-197.
- Graham, J.W., Olchowski, A.E., & Gilreath, T.D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, 8, 206-213.
- Judd, C.M., & Kenny, D.A. (1981). Process analysis: estimating mediation in treatment evaluations. *Evaluation Review*, 5, 602-619.
- Liu, L.C., Flay, B.R., et al. (2009). Evaluating Mediation in Longitudinal Multivariate Data: Mediation effects for the Aban Aya Youth Project drug prevention program. *Prevention Science*, 10, 197-207.
- MacKinnon, D.P. (2008). *Introduction to statistical mediation analysis*. Mahwah, NJ: Erlbaum.
- MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods*, 39, 384-389.
- MacKinnon, D. P., Lockwood, C. M., Brown, Wang, W., C.H., & Hoffman, J. M. (2007). The intermediate endpoint effect in logistic and probit regression. *Clinical Trials*, 4, 499-513.
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7, 83-104.

- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*, 99-128.
- MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research, 30*(1), 41-62.
- Meeker, W. Q., Cornwell, L. W., & Aroian, L. A. (1981). The product of two normally distributed random variables. In W. Kenney & R. Odeh (Eds.), *Selected tables in mathematical statistics* (Vol. VII, pp. 1-256). Providence, IR: American Mathematical Society.
- Moe, E.L., Elliot, D.L., Goldberg, L. et al. (2002). Promoting Healthy Lifestyles: Alternative Models' Effects (PHLAME). *Health Education Research, 17*, 586-596.
- Ranby, K., MacKinnon, D.P., Fairchild, A.J., Elliot, D.L., Kuehl, K.S., & Goldberg, L. (2011). The PHLAME (Promoting Healthy Lifestyles: Alternative Models' Effects) Firefighter Study: Testing Mediating Mechanisms. *Journal of Occupational Health Psychology, 16*, 501-513.
- Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. Hoboken, NJ: Wiley.
- Saxe, G.A., Major, J.M., Westerberg, L., Khandrika, S., & Downs, T.M. (2008). Biological Mediators of Effect of Diet and Stress Reduction on Prostate Cancer. *Integrative Cancer Therapies, 7*, 130-138.
- Schafer, J.L. (1997). *Analysis of incomplete multivariate data*. Boca Raton, FL: Chapman & Hall.

- Schafer, J.L., & Olsen, M.K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research*, *33*, 545-571.
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, *7*(4), 422-445.
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, *13*, 290-312.
- Tanner, M.A., & Wong, W.H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, *82*, 528-540.
- van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research*, *16*, 219-242.
- Williams, J., & MacKinnon, D. P. (2008). Resampling and distribution of the product methods for testing indirect effects in complex models. *Structural Equation Modeling*, *15*, 23-51.
- Yuan, Y., & MacKinnon, D.P. (2009). Bayesian mediation analysis. *Psychological Methods*, *14*, 301-322.

Table 1

Maximum Likelihood Descriptives from Data Analysis Example 1

	1.	2.	3.	4.	5.	6.
1. Friends' healthy eating habits	1.000	0.488	-0.009	0.333	-0.059	0.299
2. Partner's healthy eating habits	0.243	1.000	0.078	0.818	0.057	0.571
3. Treatment program (X)	-0.016	0.093	1.000	0.129	0.048	0.079
4. Coworker eating habits (M_1)	0.234	0.404	0.220	1.000	0.024	0.265
5. Knowledge of health benefits (M_2)	-0.055	0.037	0.108	0.022	1.000	0.298
6. Healthy eating intentions (Y)	0.169	0.227	0.108	0.150	0.221	1.000
Mean	3.329	3.858	0.569	3.642	6.237	2.951
Std. Dev.	1.188	1.693	0.491	1.196	0.910	1.484
% Missing	0.456	22.096	0.000	21.185	21.640	43.052

Note. Covariances are on the upper diagonal and correlations are on the lower diagonal.

Table 2

Confidence Interval Coverage from the MAR Simulation

<i>N</i>	$\alpha\beta$	$\alpha\beta$ Skew.	$\alpha\beta$ Kurt.	Bayes	MI (MVD)	ML (MVD)	ML (BCB)	ML (PB)
$\alpha = 0, \beta = 0$								
50	0	-0.009	4.265	0.999	0.999	1.000	0.993	0.997
100	0	0.012	4.079	0.998	1.000	1.000	0.992	0.997
300	0	0.009	3.970	0.999	1.000	1.000	0.993	0.998
1000	0	0.035	3.964	1.000	1.000	1.000	0.986	0.997
$\alpha = .10, \beta = .10$								
50	0.01	0.119	3.847	0.996	0.999	0.986	0.981	0.994
100	0.01	0.257	3.290	0.995	0.999	0.970	0.966	0.991
300	0.01	0.522	2.224	0.978	0.965	0.909	0.914	0.963
1000	0.01	0.612	0.894	0.950	0.928	0.920	0.954	0.939
$\alpha = .30, \beta = .30$								
50	0.09	0.559	1.854	0.952	0.934	0.900	0.932	0.943
100	0.09	0.587	0.993	0.947	0.918	0.909	0.948	0.939
300	0.09	0.437	0.322	0.951	0.936	0.936	0.954	0.947
1000	0.09	0.255	0.093	0.953	0.950	0.951	0.950	0.948
$\alpha = .50, \beta = .50$								
50	0.25	0.516	0.717	0.952	0.935	0.922	0.949	0.938
100	0.25	0.414	0.328	0.945	0.936	0.930	0.944	0.941
300	0.25	0.257	0.102	0.955	0.952	0.947	0.952	0.949
1000	0.25	0.144	0.029	0.950	0.946	0.949	0.947	0.946
$\alpha = .30, \beta = .10$								
50	0.03	0.280	2.699	0.980	0.996	0.959	0.930	0.970
100	0.03	0.377	1.747	0.968	0.978	0.947	0.917	0.956
300	0.03	0.351	0.658	0.951	0.952	0.945	0.935	0.944
1000	0.03	0.220	0.193	0.948	0.952	0.949	0.941	0.942

Note. MI = multiple imputation, ML = maximum likelihood, MVD = multivariate delta method (i.e., Sobel) test, BCB = bias corrected bootstrap, PB = percentile bootstrap.

Table 3

Empirical Power Estimates from the MAR Simulation

<i>N</i>	$\alpha\beta$	$\alpha\beta$ Skew.	$\alpha\beta$ Kurt.	Bayes	MI (MVD)	ML (MVD)	ML (BCB)	ML (PB)
$\alpha = .10, \beta = .10$								
50	0.01	0.119	3.847	0.004	0.001	0.001	0.015	0.006
100	0.01	0.257	3.290	0.012	0.001	0.002	0.032	0.015
300	0.01	0.522	2.224	0.064	0.016	0.018	0.110	0.059
1000	0.01	0.612	0.894	0.498	0.273	0.291	0.539	0.429
$\alpha = .30, \beta = .30$								
50	0.09	0.559	1.854	0.124	0.057	0.065	0.212	0.134
100	0.09	0.587	0.993	0.426	0.260	0.271	0.542	0.425
300	0.09	0.437	0.322	0.974	0.951	0.953	0.981	0.973
1000	0.09	0.255	0.093	1.000	1.000	1.000	1.000	1.000
$\alpha = .50, \beta = .50$								
50	0.25	0.516	0.717	0.666	0.595	0.563	0.744	0.645
100	0.25	0.414	0.328	0.970	0.961	0.953	0.978	0.968
300	0.25	0.257	0.102	1.000	1.000	1.000	1.000	1.000
1000	0.25	0.144	0.029	1.000	1.000	1.000	1.000	1.000
$\alpha = .30, \beta = .10$								
50	0.03	0.280	2.699	0.027	0.007	0.014	0.056	0.028
100	0.03	0.377	1.747	0.067	0.023	0.025	0.117	0.068
300	0.03	0.351	0.658	0.274	0.194	0.188	0.336	0.263
1000	0.03	0.220	0.193	0.745	0.745	0.720	0.749	0.728

Note. MI = multiple imputation, ML = maximum likelihood, MVD = multivariate delta method (i.e., Sobel) test, BCB = bias corrected bootstrap, PB = percentile bootstrap.

Figure 1. Mediation model for Data Analysis Example 1.

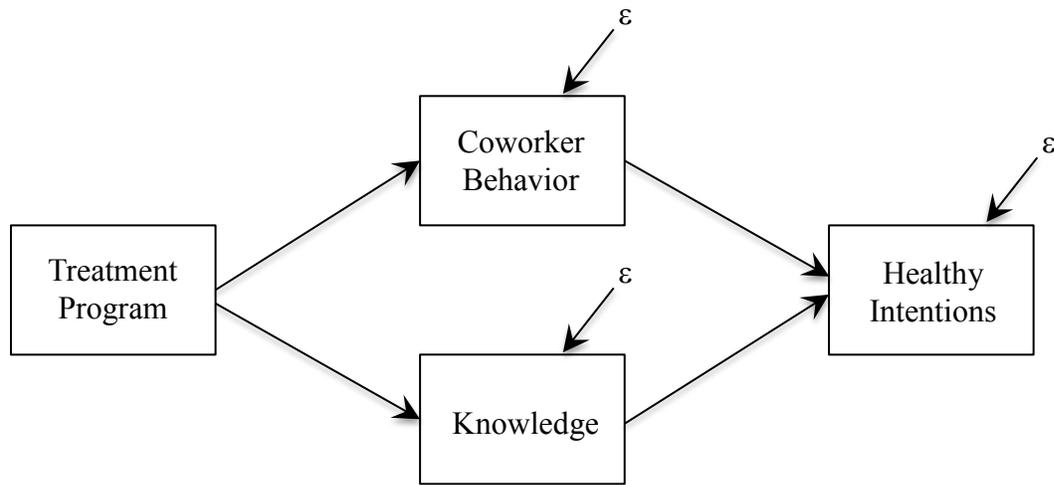


Figure 2. SAS macro input for Data Analysis Example 1. The values in bold typeface denote user-specified values.

```
%macro mediation;

/* SPECIFY INPUT DATA OPTIONS */

%let data = "c:\temp\phlame.dat";
%let varlist = txgroup coworker knowledge
intentions friends partner;
%let missingcode = -99;

/* SPECIFY PRIOR DISTRIBUTION OPTIONS */

%let prior = ;
%let priorvarlist = ;

/* SPECIFY ANALYSIS VARIABLES */

%let x = txgroup;
%let m = coworker knowledge;
%let y = intentions;
%let auxiliary = friends partner;

/* SPECIFY MCMC OPTIONS */

%let burnin = 500;
%let iterations = 10000;
%let seed = 20319;
```

Figure 3. Trace plot of the $\alpha\beta$ values from the first 500 cycles of data augmentation. The absence of long-term upward or downward trends suggests that a stable posterior distribution generated the simulated parameter values.

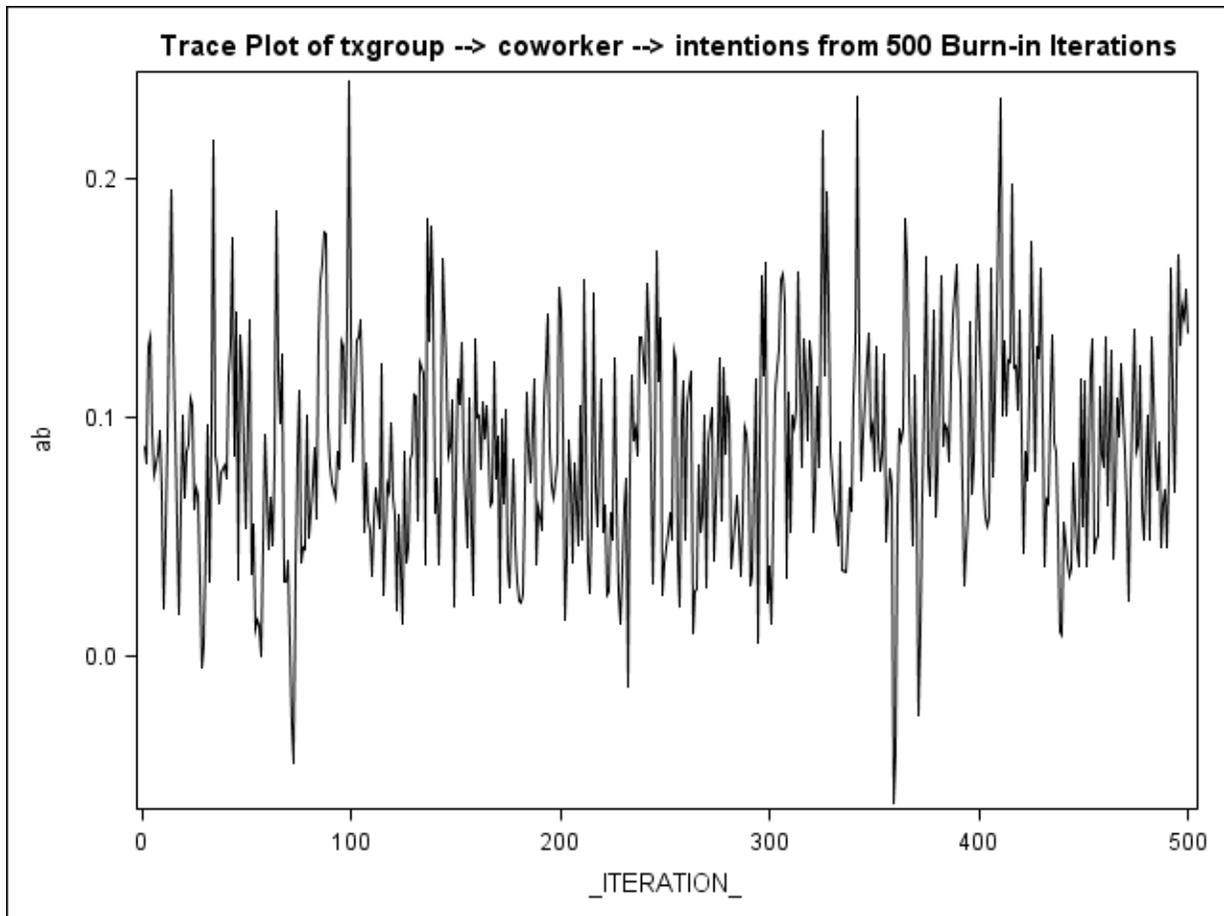


Figure 4. Kernel density plot of the $\alpha\beta$ posterior distribution from 10,000 cycles of data augmentation.

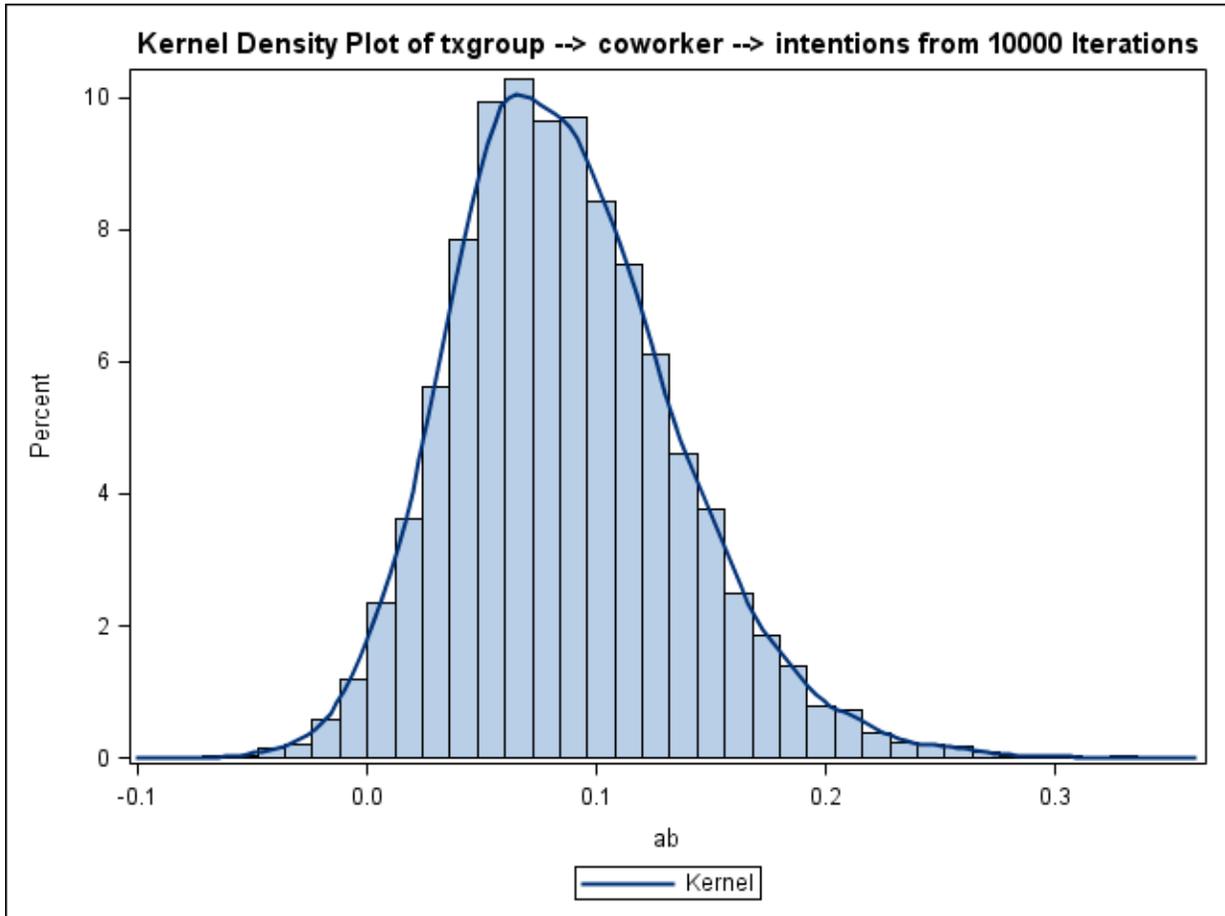


Figure 5. Macro output from Data Analysis Example 1.

SUMMARY OF POSTERIOR DISTRIBUTIONS

Total Effect of X on Y (Tau Paths)

Estimate	Median	LCL	UCL
txgroup -> intentions	0.323235	-.054293	0.696910

Direct Effect of X on M (A Paths)

Estimate	Median	LCL	UCL
txgroup -> coworker	0.533389	0.282380	0.785818
txgroup -> knowledge	0.197340	0.000523	0.394753

Direct Effect of M on Y Controlling for X (B Paths)

Estimate	Median	LCL	UCL
coworker -> intentions	0.163396	0.001897	0.323125
knowledge -> intentions	0.344401	0.143104	0.542487

Direct Effect of X on Y Controlling for M (Tau Prime Paths)

Estimate	Median	LCL	UCL
txgroup -> intentions	0.166932	-.216748	0.552133

Indirect Effect of X on Y via M (AB Paths)

Estimate	Median	LCL	UCL
txgroup -> coworker -> intentions	0.083065	0.000997	0.192888
txgroup -> knowledge -> intentions	0.063327	-.000073	0.161574

Proportion Mediated Effect Sizes

Estimate	Median	LCL	UCL
txgroup -> coworker -> intentions	0.264484	0.024072	3.770156
txgroup -> knowledge -> intentions	0.202989	0.019962	2.500836

Figure 6. SAS macro input for Data Analysis Example 2. The values in bold typeface denote user-specified values.

```
%macro mediation;

/* SPECIFY INPUT DATA OPTIONS */

%let data = "c:\temp\phlame.dat";
%let varlist = txgroup coworker knowledge
intentions friends partner;
%let missingcode = -99;

/* SPECIFY PRIOR DISTRIBUTION OPTIONS */

%let prior = "c:\temp\ex2priorcov.dat";
%let priorvarlist = _type_ $ _name_ $ txgroup coworker intentions;

/* SPECIFY ANALYSIS VARIABLES */

%let x = txgroup;
%let m = coworker;
%let y = intentions;
%let auxiliary = ;

/* SPECIFY MCMC OPTIONS */

%let burnin = 500;
%let iterations = 10000;
%let seed = 20319;
```

Figure 7. Contents of the `_TYPE_ = COV` text data set containing the prior parameters for Data Analysis Example 2.

COV	txgroup	0.2500	0.1875	0.0750
COV	coworker	0.1875	2.2500	0.4500
COV	intentions	0.0750	0.4500	2.2500
N	.	30.0000	30.0000	30.0000
MEAN	.	0.5000	3.5000	3.2500
N_MEAN	.	30.0000	30.0000	30.0000

Figure 8. Kernel density graph of the posterior distribution of $\alpha\beta$ from 10,000 cycles of parameter simulation. The dashed distribution reflects a non-informative prior, and the solid distribution is based on an informative prior with 30 degrees of freedom.

