



Sequential Bayesian Data Synthesis for Mediation and Regression Analysis

Ingrid C. Wurpts¹ · Milica Miočević² · David P. MacKinnon¹

Accepted: 26 April 2021 / Published online: 21 July 2021
© Society for Prevention Research 2021

Abstract

Science is an inherently cumulative process, and knowledge on a specific topic is organized through synthesis of findings from related studies. Meta-analysis has been the most common statistical method for synthesizing findings from multiple studies in prevention science and other fields. In recent years, Bayesian statistics have been put forth as another way to synthesize findings and have been praised for providing a natural framework for update existing knowledge with new data. This article presents a Bayesian method for cumulative science and describes a SAS macro %SBDS for synthesizing findings from multiple studies or multiple data sets from a single study using three different methods: meta-analysis using raw data, sequential Bayesian data synthesis, and a single-level analysis on pooled data. Sequential Bayesian data synthesis and Bayesian statistics in general are discussed in an accessible manner, and guidelines are provided on how researchers can use the accompanying SAS macro for synthesizing data from their own studies. Four alcohol use studies were used to demonstrate how to apply the three data synthesis methods using the SAS macro.

Keywords Data synthesis · Meta-analysis · Bayesian statistics · Sequential bayesian data synthesis

Introduction

Science is an inherently cumulative process. Just as the current paper cites multilevel mediation work by MacKinnon (2008), he cited earlier mediation work by Baron and Kenny (1986), who cited Campbell and Fiske on construct validity (1959), who cited Thurstone (1931) on reliability and validity of tests. Directly and indirectly, the research in this paper has been informed by years of prior research, going back at least 80 years. In prevention science, social influence school-based drug prevention programs have been improved since they were first introduced in the 1970s (Cuijpers, 2002; McBride, 2003; Tobler, 1997). After a body of research has

been produced on a given topic, it is important to summarize the findings from all relevant studies. Furthermore, combining findings from studies is beneficial when the sample sizes are small, and/or the base rate of a behavior being studied is low (Curran & Hussong, 2009). In prevention science, there are many examples of multiple prevention studies of different programs, including school-based health promotion and interventions for children at risk.

This paper provides instruction on performing sequential Bayesian data synthesis (SBDS) for mediation analysis, along with an accompanying SAS¹ macro, for synthesizing data across multiple studies, sites, and/or data collection periods. Using an example data set, we compare the results from SBDS with two other popular methods of data synthesis: regression analysis with pooled data and meta-analysis on raw data (multilevel analysis), which are also available in the macro. The remainder of the introduction will give a theoretical overview of the three methods for data synthesis

This research was supported in part by the National Institute on Drug Abuse (R37DA09757). At the time the majority of this research was completed, Dr. Wurpts was a graduate student at Arizona State University. She is now a Scientist at Presbyterian Healthcare Services.

✉ Ingrid C. Wurpts
icarlsol@asu.edu

¹ Arizona State University, Tempe, AZ, USA

² McGill University, Montreal, Canada

¹ We hope that by providing an annotated SAS macro, prevention science researchers and graduate students at all levels will feel comfortable performing SBDS with their own data. If an academic researcher does not have access to an institutional SAS license, SAS University edition can be downloaded for free from https://www.sas.com/en_us/software/university-edition.html.

implemented in the SAS macro. After a theoretical description of the methods, the four example data sets will be described, followed by an application of the SAS macro and a discussion of the results obtained using the three methods. The macro is annotated in a way that allows users to easily modify the macro when synthesizing their data. The paper will conclude with additional options users can implement in the macro, as well as a brief mention of related methods for data synthesis that were not implemented in this paper.

Meta-Analysis

Meta-analysis has been the most common method for summarizing findings from multiple studies in the social sciences and the prevention literature (Hedges & Olkin, 1985). Generally, meta-analysis consists of five stages as summarized in Cook et al. (1992): (1) specification of the research problem, (2) identification of relevant research studies, (3) retrieval of data such as effect sizes from research studies, (4) analysis of data from studies and interpretation of results, and (5) public presentation in a research document. When analyzing the data (stage 4), researchers have the choice between a fixed-effects model that assumes the true effect is the same in all studies, and a random-effects model that assumes the true effect varies between studies being synthesized (Brockwell & Gordon, 2001). The choice between fixed- and random-effects models depends on the type of inference desired. If the researcher is making *conditional inferences*, i.e., inferences only about the set of observed studies, then a fixed-effect model is appropriate. If the researcher is making *unconditional inferences*, i.e., using the selected sample of studies to make inferences about the population from which they are drawn, then a random-effects model is appropriate (Cooper et al., 2009).

Meta-analysis can be carried out on summary data from multiple studies as well as individual participant data (IPD). Both types of data are amenable to meta-analysis via multilevel modeling. IPD is often preferable as it contains more information than corresponding summary data (Jones et al., 2009), and meta-analysis can be performed with IPD by estimating a standard multilevel model where individual observations are nested within studies (Hox, 2002). In this paper, we use the term “raw data,” rather than IPD, as data synthesis methods are applicable to other situations where the most granular data may not be individual participants per se.

Meta-analysis is generally performed for one parameter at a time, thus making it more challenging to summarize studies with different covariates, and more cumbersome when summarizing more than one effect from multiple studies,

given that the effects must be summarized individually, rather than simultaneously (de Leeuw & Klugkist, 2012). The SAS macro accompanying this paper provides point and interval estimates of the mediated effect computed as the product of the effect of the independent variable on the mediator and the effect of the mediator on the outcome controlling for the effect of the independent variable.

Regression and Mediation Analyses on Pooled Data

Integrative data analysis (IDA) is a framework for analyzing pooled individual participant data from multiple studies (Hussong et al., 2013). IDA has numerous benefits, ranging from increases in power due to larger sample sizes than those of individual studies to increase in sample heterogeneity and frequencies of low base rate behaviors (Curran & Hussong, 2009). In IDA, individual participant data from relevant studies are first pooled, then a commensurate metric across studies is created (unless all studies used the same measurement instruments for the constructs in the model), and finally the statistical model of interest is fit using participant scores on the commensurate measures. In this project, all studies use the same measurement instruments for relevant constructs; therefore, all variables in the model already have common metrics across studies. In pooled data mediation analysis, the first step is pooling all data sets to be synthesized. In the bivariate case, the next step is to fit a regression model on the pooled data set. In the mediation case, the next step is to fit two regression models on the pooled data set: one predicting the mediator from the independent variable and the other to predict the outcome from the independent variable and the mediator. (While the approach generally can accommodate covariates, they are not currently part of the macro.) The four studies used in the empirical example are similar in design, construct measurement, and historical time (all studies were carried out within 8 years). The intraclass correlations (ICC) for M and Y were 0.0005 and 0.007, respectively, indicating that out of the total variation in M , 0.05% was due to between-study variance, and 0.7% of the total variation of Y was due to between-study variance. Although these values may appear small, even ICC estimates of 0.01 may bias estimates if the nested structure is ignored (MacKinnon, 2008; Muthén & Satorra, 1995).

However, we recognize that pooling data without accounting for between-study variation is a common technique, so we demonstrated it here for comparison purposes. In many applications, this method may be inferior to meta-analysis and SBDS because unlike the multilevel model, it does not

account for the potential dependence of observations from the same study (or site), and unlike SBDS, it does not allow for examining intermediate findings before all data sets have been synthesized. Researchers are advised to first estimate ICCs for M and Y and consider a multilevel model or SBDS approach if the values are non-trivial.

Sequential Bayesian Data Synthesis

If a researcher wants to combine information across at least two data sets, (e.g., a pilot and a main study, at least two separate studies, or data sets from the same study collected at different sites), an alternative option is sequential Bayesian data synthesis (SBDS). Methods for Bayesian data synthesis, including meta-analysis, have been described (Smith et al., 1995; Hartung et al., 2008), as well as Bayesian methods for updating the evidence for competing hypotheses (Kuiper et al., 2013) and Bayesian methods for updating linear regression findings using published summaries to construct priors (de Leeuw & Klugkist, 2012). SBDS in this context refers to a sequential updating of model parameters in studies $j=1$ to J , with J being the number of data sets to be synthesized. SBDS uses raw data and extracts point summaries from the results of the Bayesian analysis of study j to use as prior information for the Bayesian analysis of study $j+1$. We next describe the steps to perform SBDS. These steps assume that there are at least two sets of raw data which can be ordered temporally. Secondly, these steps assume that the same model can be fit in all data sets. In the example we present, this assumption is met clearly, because each data set was collected using the same survey instruments and population. However, there may be cases where the variables or methods of data collection vary among the data sets. We will discuss later how to deal with such issues with SBDS.

Sequential Bayesian data synthesis consists of 6 steps.

1. Analyze the first study with an estimation method of choice and record coefficient estimates and corresponding standard errors (or standard deviations of the posterior), sample size, and residual variances. This step could utilize frequentist methods or Bayesian analysis with diffuse or informative prior distributions.
2. Use the results from step 1 as hyperparameters for a Bayesian analysis of study 2.
3. Save posterior summaries (generally, central tendency and variation of the posterior distribution) for use as the hyperparameters for the prior in the analysis of the next study.
4. Repeat Steps 2–3 j times for each of J studies. Continue using the posterior of each study as the prior for the analysis of the subsequent study until all the studies have been analyzed.

5. Summarize the posterior distributions from the final study and draw inferences about the parameters of interest.
6. Return to Step 2 when additional future studies are available (Fig. 1). As more data are collected, the amount of prior information for each analysis increases, and so does the accuracy of the scientific knowledge summarized by the posterior distribution.

Advantages of Sequential Bayesian Data Synthesis

The SBDS approach has some unique utilities compared to other methods for data synthesis, many of which are incorporated into the %SBDS macro. SBDS accounts for temporal ordering in the accumulation of knowledge, which allows the researcher to detect after how many studies an effect

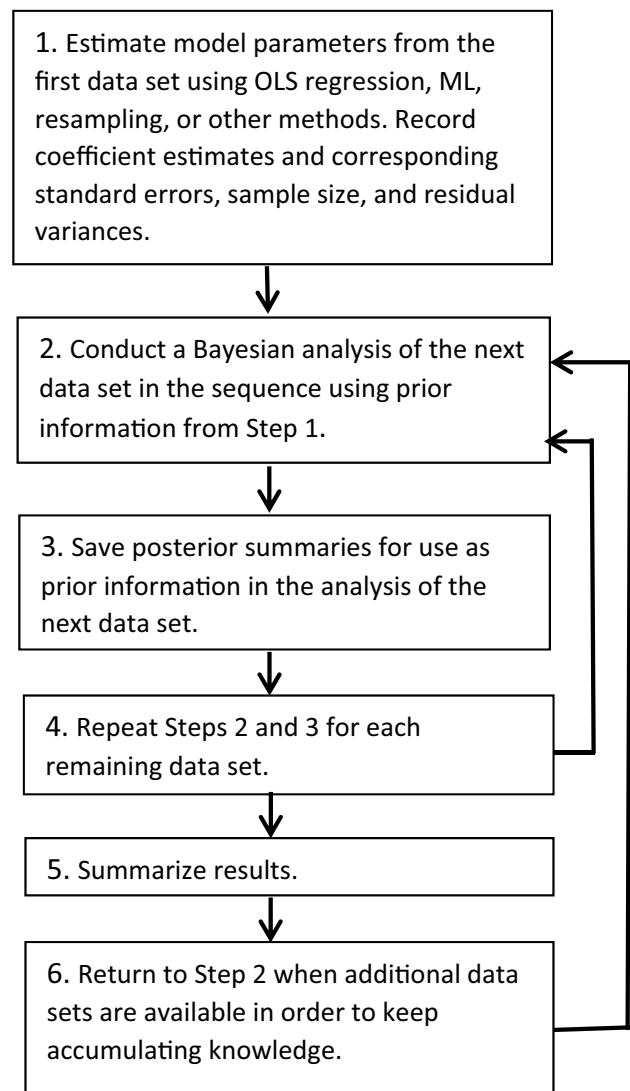


Fig. 1 Steps in a sequential Bayesian analysis

becomes significant² and whether it remains significant after additional studies have been synthesized. This method of cumulative analysis might be useful in contexts where an effect could change over time (such as school-based drug prevention programs) and the researcher is interested in observing this change as well as finding the most up-to-date effect. Although Lau et al. (1992) described a similar method of (frequentist) cumulative meta-analysis which pools the accumulated data at each step, the SBDS method more explicitly models the accumulation of knowledge via priors for each study separately. While the final step in cumulative meta-analysis reflects the pooling of all available data, the final step in SBDS estimates the effect in the final study, given prior information from all previous studies.

SBDS can also incorporate information from pilot studies and other previous studies. Whereas pilot studies are primarily used to assess the feasibility of the main study and suggest changes to the methodology prior to collecting data from the focal sample, they can also be used as relevant pieces of prior information in a statistical analysis. In the case where there is only one pilot or previous study and only one main study, multilevel methods cannot be used to combine the data, as there are only two data points used to fit the random (or fixed) effects model. Even in the case of 3, 4, or 5 studies, a multilevel model may be difficult to estimate, depending on the study size. SBDS can still be used in these cases. Doing the analysis on pooled data would be another plausible option if the raw data are available, as the feasibility of this method does not depend on how many studies are being combined.

However, as discussed previously, even small amounts of between-study variance can affect the results of a pooled data analysis. Additionally, when using data from different studies and/or sites, researchers must choose whether to consider the data collection sites and participants from different locations exchangeable, or whether their model ought to reflect these differences (e.g., in study design or population from which the sample was drawn). If some of the studies being synthesized did not sample from the population of interest for the research question, and/or used experimental manipulations that are not identical to the experimental manipulation for which researchers want to make inferences, then it may be appropriate to give such data sets less weight in the data synthesis. Downweighing the results

from such non-exchangeable studies can be accomplished in the %SBDS macro either by (1) multiplying the posterior standard deviation estimate for study j by a weight, which correspondingly decreases the influence of studies 1 through j on study $j + 1$, or (2) specifying a power prior distribution for study $j + 1$ based on the informativeness of pooling studies 1 through j (Ibrahim & Chen, 2000).

Power prior distributions are a class of informative prior distributions based on the likelihood of a historical data set raised to the power a_0 (Ibrahim & Chen, 2000). The power parameter is generally chosen so $0 \leq a_0 \leq 1$. If the likelihood function of studies 1... j is raised to the power $a_0 = 0$, then the power prior distribution for study $j + 1$ is not influenced by previous studies. On the other hand, if $a_0 = 1$, then the power prior based on studies 1... j is equivalent to using results from all observations in studies 1... j to construct an informative prior distribution for study $j + 1$. Both options will necessarily reduce the influence of studies 1 through j on study $j + 1$. The first option, however, is most useful if researchers are interested in interpreting the results at each step as weights can be specified for every additional study added to the synthesis.

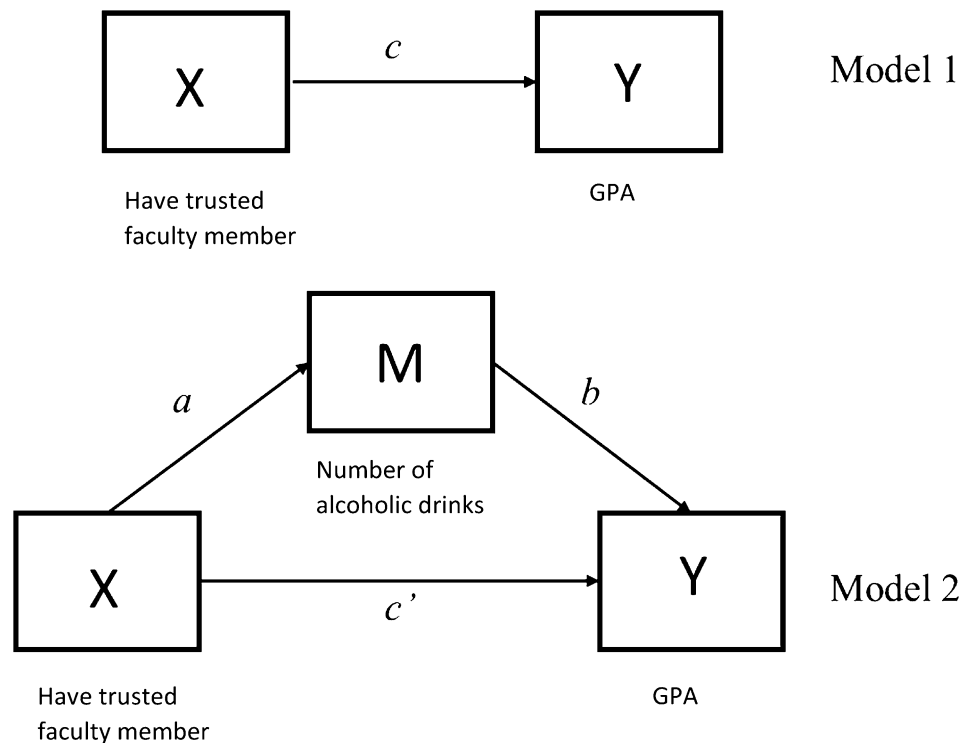
Although not currently available as a pre-programmed option in the %SBDS macro, the SBDS method can also combine raw and summary data together. First, existing prior information would be summarized via a Bayesian or frequentist meta-analysis, and the results from this analysis could be used as prior information for subsequent Bayesian analyses of raw data sets.

More generally, Bayesian analysis allows for probabilistic interpretation of effects and for computing the probability of the null hypothesis (Kruschke, 2011). In the %SBDS macro, one can compute the probability that the mediated effect is equal to any given constant, including zero; this can be done by computing the percentage of posterior draws from the posterior distribution of the last step of SBDS that lie in an interval around zero. The boundaries of the interval should be chosen such that the effect is close enough to zero so as to be non-existent, e.g., between -0.01 and 0.01 , although the range of what counts as non-existent will vary between fields and depending on the research question.

Note that we are not introducing SBDS as a superior method to multilevel meta-analysis. The two methods differ in their goals and interpretation. Whereas multilevel analysis allows estimation of an average treatment effect and the heterogeneity of that effect across studies, SBDS can increase the precision of an estimated treatment effect for a given study by considering information from other studies. In an ideal world, a researcher would have access to raw data from all relevant studies, and the properties of the data would not preclude the researcher from choosing the analysis method that is best suited to the research question. However, we recognize that is not always the case in practice. The number of studies with available raw data may be too small to

² Here we use the term significant to mean that the credibility intervals for a given effect do not contain zero. There is no concept of significance testing in the Bayesian framework, however, if 0 is not included in the credibility interval for an effect, we can conclude that the effect is different from zero. Throughout the paper we will use the term “significant” for both frequentist and Bayesian findings because of its familiar interpretation, but we caution the reader that this expression does not stem from the Bayesian theoretical framework.

Fig. 2 Bivariate model (model 1) and single mediator model (model 2)



perform multilevel analysis. Or, as we have found in our own research, attempting to account for the multilevel structure of a data set while also estimating a complex model with many parameters (such as a growth mixture model) may result in non-convergence, out-of-bounds estimates, and other errors. The SBDS method described here may be a feasible alternative to other data synthesis methods, especially for models that are more complex than OLS regression. Concepts in Bayesian statistics needed by the reader to understand the logic behind SBDS are described in Online Resource 1. This paper will proceed by describing four studies of college students' alcohol use. We use these data to demonstrate how SBDS can combine information across multiple studies and compare SBDS with raw data meta-analysis and pooled data analysis. The paper will conclude with a discussion on the usefulness and applicability of SBDS in social sciences.

Data Example

We demonstrate how to conduct SBDS with data from four rounds of the Harvard Public Health Alcohol Study (Wechsler, 1993, 1997, 1999, 2001). The study included students' self-reported use of alcohol, tobacco, and illicit drugs, as well as information related to their studies (e.g., GPA). We use these data for illustrating data synthesis methods and do not intend to draw scientific conclusions about alcohol

use and GPA. The independent variable in our mediation model represents students' answers to the question "Is there a member of the faculty or administration with whom you could discuss a problem?" (response options 1 = "Yes" and 0 = "No"). The mediator was the students' response to the question "In the past 30 days on those occasions when you drank alcohol, how many drinks did you usually have?" (response options 1–9). The outcome is student-reported GPA on a scale from 1 ("D") to 9 ("A"). Following listwise deletion for simplicity of demonstration,³ the 1993 data set had 10,528 observations, the 1997 data had 9153 observations, the 1999 data had 8887 observations, and the 2001 data set had 7118 observations. Data coding details are available in Online Resource 2, and the derived data set we used is available at https://figshare.com/articles/dataset/Harvard_alcohol_csv/12671105.

When analyzing only one data set, model 1 (Fig. 2) can be described using Eq. (1):

$$Y = i_1 + cX + e_1, \quad (1)$$

³ PROC MCMC is able to accommodate missing data via the MISSING= option, although we did not add this explicitly to the macro, as a thorough discussion of the Bayesian treatment of missing data was beyond the scope of this article. Users who wish to utilize the PROC MCMC missing data utility are referred to https://documentation.sas.com/?cdcId=pgmsascdc&cdcVersion=9.4_3.4&docsetId=statug&docsetTarget=statug_mcmc_details61.htm&locale=en

where i_1 is the intercept, c is the regression coefficient for predicting Y from X , and e_1 is the residual. Model 2 is described using Eqs. (2) and (3):

$$M = i_2 + aX + e_2 \quad (2)$$

$$Y = i_3 + c'X + bM + e_3, \quad (3)$$

where i_2 and i_3 are the intercepts in the equations predicting M and Y (respectively), a is the effect of X on M , b and c' are the conditional effects of M and X on Y (respectively), and e_2 and e_3 are the residuals from the equations predicting M and Y (respectively). The mediated effect is computed as either the product of coefficients ab or the difference of coefficients $c - c'$, and these are equivalent for the analyses conducted in this article (MacKinnon et al., 1995). This paper will use ab as a measure of the mediated effect (MacKinnon, 2008).

Three different data synthesis methods were used for both models: a SBDS of all four data sets in the order they were collected, multilevel model with a random intercept (thus, assuming a fixed slope parameter⁴ but different intercepts in the four studies) and the study number as the clustering variable, and OLS regression analysis on the pooled data from all four studies. All three data synthesis methods were computed using SAS University Edition Windows, and the %SBDS macro for performing these analyses is available in Online Resource 3. Subsequent sections will outline the procedures and results by data synthesis method.

Meta-Analysis Using Raw Data

Meta-analysis is usually performed by using point estimates, such as effect sizes, from each study as raw data are often unavailable. However, when raw data sets are available, they are more advantageous to use than summary statistics (Curran & Hussong, 2009; Jones et al., 2009). Since raw data for all four studies were available for our example, a random-intercept multilevel model was specified in SAS PROC MIXED, using study as the clustering variable. These models had four level 2 units (studies) with varying numbers of level 1 units (participants) ranging from 7118 to 10,528. The equation describing the multilevel model for the bivariate case is:

⁴ A multilevel model with random slopes was also attempted following the SAS code from Bauer, Preacher, and Gil (2006); however, this model failed to converge. Researchers who have reason to believe that the a and/or b paths (slopes) may differ among their data sets are encouraged to follow the excellent SAS documentation accompanying Bauer et al. (2006) available at <http://dbauer.web.unc.edu/publications/>.

$$Y_{ij} = i_1 + cX_{ij} + u_j + e_{1ij}, \quad (4)$$

where i_1 is the intercept, c is the regression coefficient assumed to be invariant between the four studies, u_j is the difference in the intercept of study j from the mean intercept i_1 , and e_{1ij} is the residual at the level of the individual. The effect of interest in the bivariate case is c . In the case of the single mediator model with random intercepts, there are two equations used to describe the model (Krull & MacKinnon, 1999; MacKinnon, 2008):

$$M_{ij} = i_2 + aX_{ij} + u_{Mj} + e_{2ij} \quad (5)$$

$$Y_{ij} = i_3 + c'X_{ij} + bM_{ij} + u_{Yj} + e_{3ij}, \quad (6)$$

where i_2 and i_3 are the intercepts in the equations predicting M and Y (respectively), a is the effect of X on M assumed to be invariant between the four studies, b and c' are the conditional effects of M and X on Y (respectively) assumed to be invariant between the four studies, u_{Mj} and u_{Yj} are the difference in the intercept of study j from the mean intercepts i_2 and i_3 (respectively), and e_{2ij} and e_{3ij} are the residuals at the level of the individual.⁵

In the bivariate case (model 1), the inclusion of the term u_j in the model indicated that the intercept i_1 was allowed to differ among the four studies. The slope c was constrained to be equal between studies, thus assuming a fixed-effects model for the effect of X on Y . Using a multilevel model where individual ratings were nested within studies, the maximum-likelihood estimate of having a trusted adult on GPA (coefficient c from Eq. (4)) was 0.416 with a 95% confidence interval ranging between 0.380 and 0.451.

In the mediation analysis (model 2), two multilevel models were estimated: the first with participants' self-reported number of drinks consumed in the past 30 days (M) as the outcome predicted by the presence of a trusted faculty member (X) to obtain the a coefficient. The second multilevel model used self-reported GPA (Y) as the outcome predicted from trusted faculty member (X) and participants'

⁵ Multilevel meta-analysis is more complex when mediating variables are considered because the mediation model at its simplest contains two relations, X to M and M to Y , compared to typical multilevel meta-analysis which consists of one bivariate X to Y relation. Information for a mediational multilevel meta-analysis may consist of: (1) within-study (level 1) relations for the independent variable, mediator, and dependent variable which are all collected in the same study, and (2) between-study (level 2) information which combines information on different parts of mediational relations across studies (MacKinnon, 2008). Both within-study and between-study relations can be examined in a multilevel meta-analysis, and the difference between these effects, called a contextual effect, can be examined via a significance test (Wurpts, 2016).

self-reported number of drinks consumed (M) to obtain the b coefficient. The intercepts i_2 and i_3 were allowed to be different between the four studies; however, the coefficients a , b , and c' were constrained to be equal between studies, thus assuming a fixed-effects model for the effect of X on M and the effects of X and M on Y . The mediated effect was computed as the product of coefficients ab . Using a multilevel model where individual ratings were nested within studies, the maximum-likelihood estimate of having a trusted faculty member on GPA through drinking behavior (ab) was 0.041 with a 95% confidence interval ranging between 0.034 and 0.047.

Regression Analysis Using Pooled Data

Data synthesis can also be performed by pooling data from multiple studies without accounting for study membership. To illustrate that, we pooled our four studies into one data set of size $N = 35,686$. Observations from all four studies were weighted equally, as there were no reasons to assume differences between participants in the four studies. We used the pooled data set to estimate both the single-predictor regression model (model 1) and the mediation analysis consisting of two OLS regression coefficients that would yield coefficients a and b for the product ab (model 2). Again, we performed this analysis for illustration and recommend that researchers use caution if ignoring the nested structure of the data. Using the parallel data analysis method with pooled data, the maximum-likelihood estimate of having a trusted faculty member on GPA (c) was 0.41 with a 95% confidence interval ranging between 0.374 and 0.446. The maximum-likelihood estimate of having a trusted faculty member on GPA through drinking behavior (ab) was 0.041 with a 95% confidence interval ranging between 0.034 and 0.047.

Sequential Bayesian Data Synthesis

The SBDS for the four data sets was carried out in the order in which the data were collected. The first data set was analyzed using a diffuse prior, and the resulting parameter estimates from this analysis were used to specify prior distributions for the Bayesian regression analysis of the second data set. Point summaries of regression coefficients (and intercepts) from the posterior distribution of the first Bayesian mediation analysis were used as mean hyperparameters for normal priors for the corresponding regression coefficients (and intercepts) in a Bayesian analysis of the second data set. The posterior standard deviations for regression coefficients (and intercepts) of the first data set

were used as standard deviation hyperparameters⁶ of the normal priors for the corresponding regression coefficients (and intercepts) in a Bayesian analysis of the second data set.

The diffuse priors used for the analysis of the first study (1993) were the normal prior with the mean hyperparameter equal to 0 and the variance hyperparameter equal to 100 times the variance of the outcome variable for regression coefficients. The inverse-gamma priors for residual variances had shape and scale hyperparameters equal to 0.5, thus encoding the assumption that the best guess for the residual variance was 1 and that the prior information about this parameter had the weight of 1 observation (Gelman et al., 2004). The priors for residual variances in the analysis of the second study (1997) were specified as inverse-gamma priors with the shape parameter equal to half of the observed sample size in the 1993 study, and the scale parameter equal to the product of the sample size and the observed residual variance in the 1993 study divided by 2. This choice of hyperparameters ensured that the bulk of the density of the prior was around the observed residual variance in study 1 (Gelman et al., 2004).

Point summaries of central tendency and variability from the posterior distributions of all 1997 study model parameters were then used as hyperparameters for the prior distributions for a Bayesian analysis of the 1999 data. For this example, we used the posterior median of all regression coefficients and intercepts as the mean of the prior distribution for the subsequent study. We performed the same analysis using the posterior mean of the regression coefficients as the mean hyperparameter of normal priors for the next study, and we found that the results did not change appreciably. Likewise, we used the standard deviation of the posterior distribution of the regression coefficients in one study as the standard deviation of the prior distribution for the same regression coefficient in the subsequent study. Priors for the residual variances of each study were specified as inverse-gamma priors with the shape parameter equal to half of the observed sample size in the preceding study, and the scale parameter equal to the product of the sample size of the preceding study, and the posterior median of the corresponding residual variance in the preceding study divided by 2. This process was repeated with the results

⁶ Note that another way to specify the spread hyperparameter of a normal prior in SAS is as a variance, which one can compute by squaring the standard errors or posterior standard deviations of the corresponding regression coefficient or intercept. Specifying a standard deviation hyperparameter to be equal to the standard deviation of the coefficient and as the variance hyperparameter to be equal to the squared standard deviation of the coefficient are just two ways of specifying the same prior distribution in SAS PROC MCMC (Miočević & MacKinnon, 2014; SAS Institute Inc., 2013).

Table 1 Posterior distribution summaries for bivariate and mediation models

Dataset	Sample Size	Parameter	Posterior Median	Posterior SD	Credible Lower	Credible Upper	HPD Lower	HPD Upper
Bivariate Model								
1993	10528	i_1	5.060	0.025	5.013	5.112	5.017	5.112
1993	10528	c	0.429	0.032	0.360	0.487	0.375	0.501
1993	10528	$e21$	3.052	0.041	2.967	3.130	2.964	3.128
1997	9153	i_1	5.105	0.018	5.073	5.139	5.075	5.140
1997	9153	c	0.446	0.025	0.387	0.488	0.402	0.495
1997	9153	$e21$	3.043	0.030	2.984	3.100	2.983	3.097
1999	8887	i_1	5.159	0.014	5.136	5.187	5.136	5.187
1999	8887	c	0.476	0.020	0.436	0.520	0.433	0.511
1999	8887	$e21$	2.923	0.030	2.864	2.981	2.864	2.981
2001	7118	i_1	5.220	0.013	5.192	5.246	5.193	5.246
2001	7118	c	0.524	0.017	0.488	0.560	0.485	0.551
2001	7118	$e21$	2.870	0.032	2.805	2.935	2.805	2.933
Mediation Model								
1993	10528	a	-0.390	0.042	-0.475	-0.306	-0.465	-0.301
1993	10528	b	-0.142	0.007	-0.156	-0.128	-0.156	-0.128
1993	10528	ab	0.055	0.007	0.043	0.069	0.042	0.068
1997	9153	a	-0.315	0.027	-0.367	-0.261	-0.370	-0.266
1997	9153	b	-0.138	0.005	-0.148	-0.128	-0.148	-0.129
1997	9153	ab	0.043	0.004	0.036	0.052	0.036	0.052
1999	8887	a	-0.275	0.022	-0.315	-0.231	-0.315	-0.233
1999	8887	b	-0.127	0.004	-0.134	-0.120	-0.135	-0.121
1999	8887	ab	0.035	0.003	0.029	0.041	0.029	0.041
2001	7118	a	-0.299	0.019	-0.337	-0.264	-0.338	-0.265
2001	7118	b	-0.118	0.003	-0.124	-0.112	-0.124	-0.112
2001	7118	ab	0.035	0.002	0.031	0.040	0.030	0.040

from 1997 becoming the prior information for the 1999 study, and so on, until the final study (2001) was analyzed. Detailed explanations of the SAS PROC MCMC syntax for this analysis can be found in the supplementary document available at URL.

For the effect of having a trusted faculty member on GPA (c), the posterior median from the initial 1993 data set ($N = 10,528$) was 0.435, posterior $SD = 0.037$, so those with a trusted faculty member had an average GPA 0.435 higher (on a 9-point scale) than those who did not have a trusted faculty member. The 95% HPD credibility interval for this effect ranged between 0.365 and 0.494. The posterior median of the intercept, i.e., the predicted GPA for those who did not have a trusted faculty member, was 5.05, posterior $SD = 0.273$. The posterior median of the residual variance for this model was 3.05.

Next, a Bayesian regression analysis of Y on X for the 1997 data ($N = 9153$) was performed using PROC MCMC. The c parameter was given a normal prior with mean of 0.435 and standard deviation of 0.037. The intercept was assigned a normal prior with a mean of 5.05 and a standard

deviation of 0.027. The variance of the residual for the analysis was given an inverse-gamma prior with a shape of $10,528/2 = 5264$ and a scale of $(3.05 * 10,528)/2 = 16,045.9$.

This analysis yielded a posterior distribution for c with a median of 0.453, $SD = 0.022$. The posterior median for the intercept was 5.10 and the posterior $SD = 0.015$. The posterior distribution for the residual variance had a median of 2.87. These posterior summaries were then used as prior information in a Bayesian regression analysis for the third data set (1999). The c parameter was assigned a normal prior with $M = 0.453$, $SD = 0.021$, and the intercept was given a normal prior with $M = 5.10$, $SD = 0.015$. The residual variance was given an inverse-gamma prior with a shape hyperparameter of $9153/2 = 4576.5$ and a scale hyperparameter of $(3.04 * 9153)/2 = 13,914$. This process of using the central tendency and variability of the posterior distributions for each analysis as the hyperparameters of the prior distribution for the next data set was continued until the final data set analyzed.

Using SBDS to combine information across all four studies, the median of the posterior distribution for the c effect in the final data set was 0.518 and the 95% HPD credibility

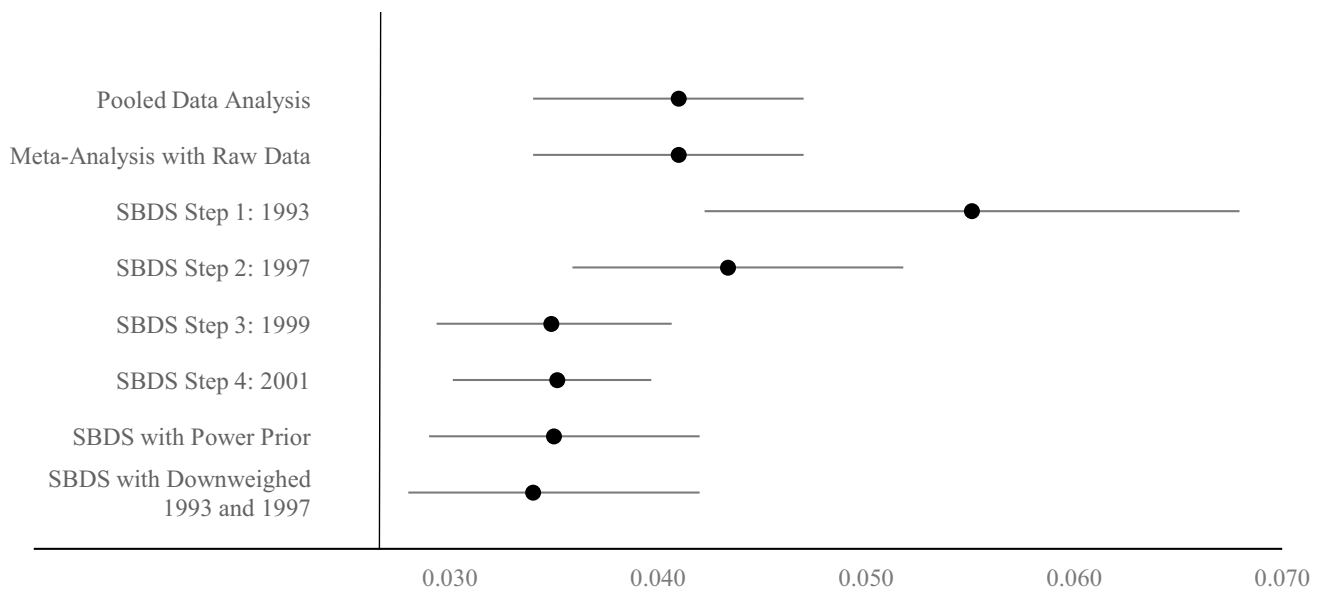


Fig. 3 Point and interval estimates of the ab effect from all analyses. For frequentist analyses (Pooled Data Analysis and Meta-Analysis), error bars represent 95% confidence intervals. For the remaining anal-

yses, point estimates are taken from the posterior median and error bars represent 95% HPD intervals

interval ranged between 0.489 and 0.542, meaning there was a 95% probability that having a trusted faculty member increased GPA between 0.489 and 0.542 units (on a 9-point scale). Table 1 includes posterior medians, posterior standard deviations, 95% equal-tail, and highest posterior density (HPD) intervals for the effect of having a trusted faculty member on GPA (c), the intercept (i_1), and the residual variance σ_{e1}^2 for all data sets.

For the mediated effect of having a trusted faculty member on GPA through alcohol consumption (ab), the posterior distribution median for the initial 1993 data set was 0.055. The 95% HPD credibility interval obtained via the %POSTINT macro (SAS Institute Inc., 2013) ranged from 0.042 to 0.068 for ab in the 1993 data set. Using SBDS with SAS PROC MCMC, the median of the posterior distribution for the mediated effect in the final data set (2001) was 0.035 with a 95% HPD credibility interval between 0.030 and 0.040. This means there was a 95% chance that having a trusted faculty member increased GPA via reduced drinking behavior between 0.03 and 0.04 units (on a 9-point scale). Table 1 includes posterior medians, posterior standard deviations, 95% equal-tail credibility intervals, and 95% highest posterior density intervals for the effect of condition on imagery (a), the effect of imagery on number of words recalled (b), and the mediated effect (ab) for all data sets analyzed using PROC MCMC.

As mentioned earlier, data sets available for synthesis may not always be homogenous in terms of quality or design. SBDS is very adaptable to such problems, as the informativeness or precision of prior information can be altered at

each step via specification of a vector of weights for each study, between 0 and 1. Studies with weight = 0 are removed from the analysis. For those with weights > 0, the inverse of each weight is used to multiply the standard deviation of the posterior distributions from that study before those values are used as hyperparameters for the subsequent study. A smaller weight, therefore, reduces the influence that a particular study has on the analysis of the subsequent study. For weights = 1, the standard deviation of the posterior distribution remains unchanged.

For demonstration purposes, if we set the weights for studies 1993 and 1997 to 0.5, the posterior standard deviations for each parameter in the 1993 analysis would be multiplied by $1/0.5 = 2$ before being used as prior distributions for the 1997 study. Likewise, the posterior standard deviations would be doubled from 1997 before being used as priors for the 1999 study. In this case, the median of the posterior distribution for the mediated effect in the final data set would be 0.034 with a 95% HPD credibility interval from 0.028 to 0.042.

Sequential Bayesian Data Synthesis with a Power Prior

The code in Online Resource 3 also allows for users to perform a semi-sequential Bayesian data synthesis where the initial study or studies are used to create a power prior for the next study, and then, the sequential Bayesian updating continues from that point. In this case, the initial priors for error terms are inverse-gamma with hyperparameters

(0.001, 0.001). For example, we used the 1993–1997 studies combined with $a_0 = 0.5$ to create a power prior for the 1999 study. In this case, the mediated effect for the 1999 study with a power prior was 0.036, with 95% HPD credibility interval between 0.026 and 0.044. The point summaries from this analysis were then used to define the priors for an analysis of the 2001 study, like described in the preceding sequential Bayesian analysis section. Using this method, the mediated effect for the final 2001 study was 0.035 with a 95% HPD credibility interval between 0.029 and 0.042.

The code also allows specification of a threshold for the mediated effect and assessment of the probability that the mediated effect is greater or less than this threshold. In our example, we set the threshold to 0 and tested the hypothesis that $ab > 0$. For all four studies, we found that 100% of the posterior draws were > 0 . If users are interested in where the HPD credibility intervals for ab exclude zero, the output includes a table with this information (all studies in our example had HPD credibility intervals which excluded 0). In addition, users can suppress plots for PROC MCMC and PROC REG, and they can specify the number of burn-in iterations, number of MCMC iterations, thinning parameter, and seed. When using the macro, it is important that the data be sorted by the order in which the user wishes to analyze the data. Across all three methods of data synthesis, final point and interval estimates (i.e., point and interval summaries in the Bayesian framework) are similar, as seen in Fig. 3.

Discussion

Using Bayesian methods to sequentially update findings in data synthesis has been advocated before (de Leeuw & Klugkist, 2012; Kuiper et al., 2013). However, to our knowledge, this is the first study to implement it in a mediation model using raw data and compare the results it produces to those obtained from more commonly used data synthesis methods, i.e., multilevel meta-analysis and OLS regression on pooled data via a SAS macro. We note that the point estimates for c and ab did not differ drastically among the three methods in our example, although the credibility intervals obtained in the SBDS were narrower (and thus more precise) than the confidence intervals from the other methods. This similarity among the different methods is unsurprising given the very large sample sizes and relative homogeneity between each study. With smaller and more heterogeneous studies, the SBDS approach offers more flexibility in terms of weighting the information from each study—in cases where some data sets are less reliable, the variance hyperparameters of the

prior distributions can be easily modified via weights to reflect this heterogeneity. Without a priori knowledge of heterogeneity, each study could be analyzed separately to ensure that the specified model is appropriate before combining into a sequential analysis. Future iterations of the %SBDS macro might include an automating of the choice of appropriate prior distributions in cases where normality is not expected.

Also, SBDS allows for the accumulation of scientific knowledge. In our example, we demonstrated how to model the accumulation of knowledge across time for the effect of having a trusted faculty advisor on college students' GPA. Again, we wish to emphasize here that our empirical example was chosen for the purposes of illustrating the SBDS method, and we assumed a temporal ordering of the three constructs. With observational data such as these, it is also possible that other relations among the data could explain our results, such as if GPA was a moderator rather than mediator, if GPA was a confounder rather than a mediator, or if all three variables measured the same construct. The SBDS method describes could be extended to other more complex mediation models (MacKinnon, 2008). Furthermore, we did not explicitly account for the non-normal distribution of the outcome variable, although mediation models for such zero-inflated distributions have been described elsewhere (O'Rourke & Vasquez, 2019).

Another important distinction between SBDS and the other two methods lies in the interpretation of the findings: while meta-analysis and OLS regression gave one point estimate of the effects of interest and intervals interpreted in terms of confidence (i.e., upon repeated sampling, 95% of the intervals constructed using these two methods will contain the true value of the effect of interest), SBDS gave an entire distribution for the parameter of interest based on which intervals with a probabilistic interpretation were formed. Thus, when using SBDS, one has the advantage of being able to conclude that the true value of c lies between 0.489 and 0.542, and the true value of ab lies between 0.028 and 0.042 with 95% probability.

Replication issues in psychology right now call for a more nuanced understanding of the accumulation of research beyond whether each study independently showed a significant or non-significant effect (Maxwell et al., 2015). Although all four alcohol studies had similarly sized effects for c and ab parameters (likely because of the large sample sizes), it is possible that a subsequent study may find parameters that are larger, smaller, or even in a different direction. Rather than viewing this hypothetical result as contrary to the results from the other data sets, SBDS allows integration of replicated studies to provide a more general understanding of the experimental effect. This facilitates the accumulation of knowledge across multiple studies even when the studies have differing results.

Note that all data synthesis methods may be prone to publication bias when the researcher is required to compile scientific findings without access to non-statistically significant findings. Studies and data sets with small effect sizes and non-significant results are often not published, and then not included in the data synthesis. Thus, studies with significant findings are likely to be overrepresented in meta-analytic literature compared to comparable studies with non-significant findings. Like many others in our field, we support the increased use of publicly available data sets to combat this so-called file drawer problem (Nosek et al., 2012).

When raw data are available, however, SBDS offers an intuitive and easily interpretable method to obtain point and interval summaries of an effect of interest across multiple studies. SBDS could also be expanded to more complicated structural and latent variable models if prior distributions for model parameters can be accurately represented in the Bayesian framework. There has been considerable growth in the amount and quality of data collected in prevention science research since the formation of the field of prevention science. The sequential Bayesian framework provides an organized way to accumulate scientific knowledge in prevention.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11121-021-01256-1>.

Funding Marie Skłodowska-Curie Individual Fellowship awarded to Milica Miočević (European Commission Horizon 2020 research and innovation program, grant number 792119). This research was supported in part by the National Institute on Drug Abuse (R37DA09757).

Declarations

Ethics Approval Not applicable. This study performed secondary data analysis on publicly available de-identified data which does not constitute human subjects research.

Consent to Participate Not applicable. This study performed secondary data analysis on publicly available de-identified data which does not constitute human subjects research.

Conflict of Interest The authors declare no competing interests.

References

- 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
- Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, *11*, 142–163
- Brockwell, S. E., & Gordon, I. R. (2001). A comparison of statistical methods for meta-analysis. *Statistics in Medicine*, *20*, 825–840
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81
- Cook, T. D., Cooper, H., Cordray, D. S., Hartmann, H., Hedges, L. V., & Light, R. J. (Eds.). (1992). *Meta-analysis for explanation: A casebook*. Russell Sage Foundation.
- Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation.
- Cuijpers, P. (2002). Effective ingredients of school-based drug prevention programs: A systematic review. *Addictive Behaviors*, *27*, 1009–1023
- Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, *14*, 81–100
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis*. CRC Press.
- Hartung, J., Knapp, G., & Sinha, B. K. (2008). Bayesian meta-analysis. *Statistical Meta-Analysis with Applications*, 155–170.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.
- Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Lawrence Erlbaum.
- Hussong, A. M., Curran, P. J., & Bauer, D. J. (2013). Integrative data analysis in clinical psychology research. *Annual review of clinical psychology*, *9*, 61–89
- Ibrahim, J. G., & Chen, M. H. (2000). Power prior distributions for regression models. *Statistical Science*, *15*, 46–60
- Institute Inc, S. A. S. (2013). *SAS/STAT® 13.1 User's Guide*. Cary, NC: SAS Institute Inc.
- Jones, A. P., Riley, R. D., Williamson, P. R., & Whitehead, A. (2009). Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials. *Clinical Trials*, *6*, 16–27
- Krull, J. L., & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review*, *23*, 418–444
- Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science*, *6*, 299–312
- Kuiper, R. M., Buskens, V., Raub, W., & Hoijtink, H. (2013). Combining statistical evidence from several studies a method using Bayesian updating and an example from research on trust problems in social and economic exchange. *Sociological Methods & Research*, *42*, 60–81
- Lau, J., Antman, E. M., Jimenez-Silva, J., Kupelnick, B., Mosteller, F., & Chalmers, T. C. (1992). Cumulative meta-analysis of therapeutic trials for myocardial infarction. *New England Journal of Medicine*, *327*, 248–254
- de Leeuw, C., & Klugkist, I. (2012). Augmenting data with published results in Bayesian linear regression. *Multivariate Behavioral Research*, *47*, 369–391
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Lawrence Erlbaum Associates.
- MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research*, *30*, 41–62
- Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist*, *70*, 487–498
- McBride, N. (2003). A systematic review of school drug education. *Health Education Research*, *18*, 729–742
- Miočević, M., & MacKinnon, D. P. (2014). SAS® for Bayesian Mediation Analysis. In Proceedings of the 39th annual meeting of SAS Users Group International. Cary, NC: SAS Institute, Inc.
- Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology*, 267–316.

- Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, 7(6), 615–631.
- O'Rourke, H. P., & Vazquez, E. (2019). Mediation analysis with zero-inflated substance use outcomes: Challenges and recommendations. *Addictive Behaviors*, 94, 16–25
- Smith, T. C., Spiegelhalter, D. J., & Thomas, A. (1995). Bayesian approaches to random-effects meta-analysis: A comparative study. *Statistics in Medicine*, 14, 2685–2699
- Thurstone, L. L. (1931). The reliability and validity of tests.
- Tobler, N. (1997). Meta analysis of adolescent drug prevention programs: Results of the 1993 meta analysis. In W. Bukoski (Ed.), *Meta-analysis of Drug Abuse Prevention Programs*. (pp. 5–68). NIDA.
- Wechsler, H. (1993). Harvard school of public health college alcohol study, 1993. ICPSR6577-v3. Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributor], 2020-01-30. <https://doi.org/10.3886/ICPSR06577.v4>
- Wechsler, H. (1997). Harvard school of public health college alcohol study, 1997. ICPSR3163-v3. Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributor], 2020-01-30. <https://doi.org/10.3886/ICPSR03163.v4>
- Wechsler, H. (1999). Harvard school of public health college alcohol study, 1999. ICPSR3818-v2. Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributor], 2020-01-30. <https://doi.org/10.3886/ICPSR03818.v3>
- Wechsler, H. (2001). Harvard school of public health college alcohol study, 2001. ICPSR04291-v2. Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributor], 2008-02-05. <https://doi.org/10.3886/ICPSR04291.v2>
- Wurpts, I. C. (2016). Performance of contextual multilevel models for comparing between-person and within-person effects (Doctoral dissertation). Retrieved from ASU Library Digital Repository.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.