Routledge
Taylor & Francis Group

Health Psychology Review

ISSN: 1743-7199 (Print) 1743-7202 (Online) Journal homepage: http://www.tandfonline.com/loi/rhpr20

An introduction to Bayesian statistics in health psychology

Sarah Depaoli, Holly M. Rus, James P. Clifton, Rens van de Schoot & Jitske Tiemensma

To cite this article: Sarah Depaoli, Holly M. Rus, James P. Clifton, Rens van de Schoot & Jitske Tiemensma (2017) An introduction to Bayesian statistics in health psychology, Health Psychology Review, 11:3, 248-264, DOI: 10.1080/17437199.2017.1343676

To link to this article: https://doi.org/10.1080/17437199.2017.1343676

Accepted author version posted online: 20 Jun 2017.
Published online: 11 Jul 2017.

HEALTH PSYCHOLOGY REVIEW, 2017
VOL. 11, NO. 3, 248–264
https://doi.org/10.1080/17437199.2017.1343676

Routledge
Taylor & Francis Group

An introduction to Bayesian statistics in health psychology

Sarah Depaoli[a], Holly M. Rus[a], James P. Clifton[a], Rens van de Schoot[b,c] and Jitske Tiemensma[a]

[a]Department of Psychological Sciences, University of California, Merced, Merced, CA, USA; [b]Department of Methods and Statistics, Utrecht University, Utrecht, The Netherlands; [c]Faculty of Humanities, Optentia Research Programme, North West University, Potchefstroom (North West), South Africa

ABSTRACT

The aim of the current article is to provide a brief introduction to Bayesian statistics within the field of health psychology. Bayesian methods are increasing in prevalence in applied fields, and they have been shown in simulation research to improve the estimation accuracy of structural equation models, latent growth curve (and mixture) models, and hierarchical linear models. Likewise, Bayesian methods can be used with small sample sizes since they do not rely on large sample theory. In this article, we discuss several important components of Bayesian statistics as they relate to health-based inquiries. We discuss the incorporation and impact of prior knowledge into the estimation process and the different components of the analysis that should be reported in an article. We present an example implementing Bayesian estimation in the context of blood pressure changes after participants experienced an acute stressor. We conclude with final thoughts on the implementation of Bayesian statistics in health psychology, including suggestions for reviewing Bayesian manuscripts and grant proposals. We have also included an extensive amount of online supplementary material to complement the content presented here, including Bayesian examples using many different software programmes and an extensive sensitivity analysis examining the impact of priors.

ARTICLE HISTORY
Received 23 May 2015
Accepted 31 August 2016

KEYWORDS
Bayesian statistics; prior distributions; convergence; posterior

Most researchers in the social and behavioural sciences will have heard of Bayesian statistics given the increasing prevalence of Bayesian estimation in these areas. A systematic review found the proportion of Bayesian articles published across fields per year has approximately tripled since 1990 (van de Schoot, Winter, Zondervan-Zwijnenburg, Ryan, & Depaoli, 2017). The current article acts as a brief introduction for using these techniques for health-related inquiries. We have also provided online supplementary material at: https://osf.io/gsr58/. Material includes: (1) syntax for many different Bayesian statistical programmes, (2) additional figures of results, (3) a convergence checking guide, (4) a thorough example of implementing Bayesian methods, and (5) a sensitivity analysis example.

A brief introduction to Bayesian estimation

We aim to provide an accessible introduction to Bayesian methods here. If the reader is interested in additional readings, then we suggest: the online material, Bolstad (2007), Hoff (2009), Kaplan and Depaoli (2013), Kruschke (2013), or van de Schoot et al. (2014). For more detailed readings, see Carlin and Louis (2009), Gelman et al. (2013), or Kruschke (2015).

CONTACT Sarah Depaoli sdepaoli@ucmerced.edu
Supplemental data for this article can be accessed at https://doi.org/10.1080/17437199.2017.1343676

Using Bayesian methods when traditional approaches fail in health psychology

Health psychologists may find Bayesian methods to be particularly helpful for health-related research inquiries. First, Bayesian methods can improve estimation accuracy (i.e., obtaining accurate parameter estimates) of complex statistical models (e.g., structural equation models, latent growth models, and multilevel models); for methodological examples showing the superior performance of Bayesian methods, see: Depaoli (2013), Hox, van de Schoot, and Matthijsse (2012), and Kim, Suh, Kim, Albanese, and Langer (2013). Second, Bayesian methods have been shown to provide more accurate results (i.e., parameter estimates) than traditional (frequentist) methods under cases of small sample sizes in a variety of different types of models (see, e.g., Depaoli & Clifton, 2015; Lambert, Sutton, Burton, Abrams, & Jones, 2005; van de Schoot, Broere, Perryck, Zondervan-Zwijnenburg, & van Loey, 2015; Zhang, Hamagami, Wang, Grimm, & Nesselroade, 2007).[1] Finally, Bayesian methods are also advantageous in cases where relatively 'simple' models are implemented (e.g., the *t*-test; Kruschke, 2013), since much richer results can be obtained compared to traditional null hypothesis statistical testing.

Although Bayesian methods have been shown to have advantages over frequentist methods, some of the same issues or problems exist across both estimation frameworks. Specifically, researchers still need to be concerned with issues of generalisability (e.g., through proper sampling techniques) and replication. With the recent recommendation to increase replication in psychology (Asendorpf et al., 2013), we echo that this is an important topic within Bayesian statistics as well.

Key 'ingredients' of the Bayesian framework: models, parameters, and probability distributions

At the heart of inferential statistics is the notion of a statistical model. Statistical models can range from rather simple to exceedingly complex, but they all have certain features in common that embody the core of statistical inference. Statistical models are often used to help capture relationships, or underlying phenomena, in the population. Specifically, they consist of probability distributions that are used to model patterns or relationships in the population among variables. In general, inferential statistics use sample data to reflect processes in the population, and to some degree the statistical model can be considered as a representation (or theory) of the data-generating process.

Statistical models are formed by mathematical equations that link variables together in a particular manner. In the case of a simple prediction model, there might be interest in the ability for an acute stressor (the model input or *predictor*) to predict changes in blood pressure (the model output or *outcome*). These variables can be linked together using a simple regression model, where a model parameter (the regression weight or coefficient) is used to show the degree to which these two variables are linked. Then the sample data are used to estimate the relationship between the acute stressor and blood pressure by estimating the value of the model parameter (i.e., the regression weight). A mathematical model (i.e., the regression model) formally captures this relationship with unknown model parameters (i.e., the regression weight for the acute stressor predictor variable).

Before collecting sample data and examining the relationship between the acute stressor and blood pressure, there might be pre-existing opinions or beliefs about this relationship. These beliefs are classified as prior beliefs because they occur previous to observing data patterns. After data collection, the magnitude of the relationship between the two variables is examined by estimating the regression weight using the sample data. Once results are obtained, beliefs about the relationship between the two variables are updated. These beliefs are classified as posterior beliefs because they occur after data collection and analysis. The next section details how Bayesian methods can take us from the notion of prior beliefs to posterior beliefs using a mathematical theorem resulting in Bayes' rule. In this context, the prior beliefs and the posterior beliefs are captured in terms of probability distributions.

Bayes' rule: a method for combining data and previous knowledge of model parameters

Let us take another example and assume two groups of individuals are compared on systolic blood pressure levels using a two-sample *t*-test. Sample data from these two groups are used to estimate their respective systolic blood pressure in the population. Then the *t*-test is used to make a comparison across these groups. It may be that Group 1 would be assumed (based on pilot studies, past literature, expert opinions, etc.) to have a mean systolic blood pressure value of 120 in the population. This belief would indicate that values near 120 would be more likely to be observed in the Group 1 sample data than values, say, near 1000 (an impossible value for this variable). Bayesian methods can be used in this situation to redistribute the likelihood (or possibility) of certain values being observed in the population by accounting for the belief that some outcomes are more likely to occur than others. In other words, Bayesian methods can incorporate the idea that values closer to 120 are more likely (i.e., more probable) compared to extreme mean values (e.g., 1000).

The process used to account for certain parameter values being more likely compared to others is through Bayes' rule. This formula contains conditional probabilities surrounding the model parameters (e.g., Group 1 population mean systolic blood pressure value), the observed sample data, and the possibility of certain population values occurring for a model parameter (e.g., systolic blood pressure for Group 1). The reduced version of Bayes' rule can be written as the following (with the symbol $\propto$ representing 'proportional to'):

$$p(\theta|y) \propto (y|\theta)p(\theta), \tag{1}$$

where θ represents all model parameters (e.g., Group 1 population mean for systolic blood pressure), and y represents the observed sample data. This equation is read as follows. The term $p(\theta|y)$ represents a conditional probability which means that the probability of the model parameters (θ) is computed given (or conditional upon) the data (y), and this term is also known as the *posterior* – this is the final result you are trying to obtain through estimation (i.e., the parameter estimate). The term $p(y|\theta)$ represents the conditional probability of the data given the model parameters, and this term represents the *data likelihood* (or sample data). The term $p(\theta)$ represents the probability of particular model parameter values existing in the population. This term is called a *prior* and, in the case of our *t*-test example, it would incorporate the idea that some mean systolic blood pressure values (e.g., 120) are more probable than others (e.g., 1000). The term *prior* might seem odd at first, but it is named so because it represents our previous (or prior) knowledge about the likely values for a given model parameter.

A recap of terminology used for Bayes' rule. Within the context of Bayesian statistics, prior knowledge or information about a model parameter can be directly incorporated into the modelling process. This prior knowledge is captured by a *prior distribution* (or *prior*).[2] Prior distributions can play a big role in Bayesian estimation. They represent hypotheses expressed as distributions over model parameters, with central tendency and uncertainty that captures the researcher's beliefs about the population parameters. Priors are combined with the data *likelihood* to form what is referred to as the *posterior*, which is the final estimate (or answer) you are trying to obtain for each model parameter.

Priors range from very informative to non-informative, depending on the amount of uncertainty (i.e., the lack of information) they contain about population parameters. Figure 1 shows the relationship between normal priors representing different levels of uncertainty (i.e., containing different levels of background knowledge) and the final posterior obtained that is used for deriving parameter estimates. A more informative prior has a greater deal of certainty surrounding the possible values a model parameter can take on, and thus it has a larger impact on the posterior (or final result).[3] In Figure 1, the first column illustrates a prior with a great deal of certainty; notice that this prior distribution has probability mass hovering over a relatively smaller range of the possible parameter space (captured through the *x*-axis).

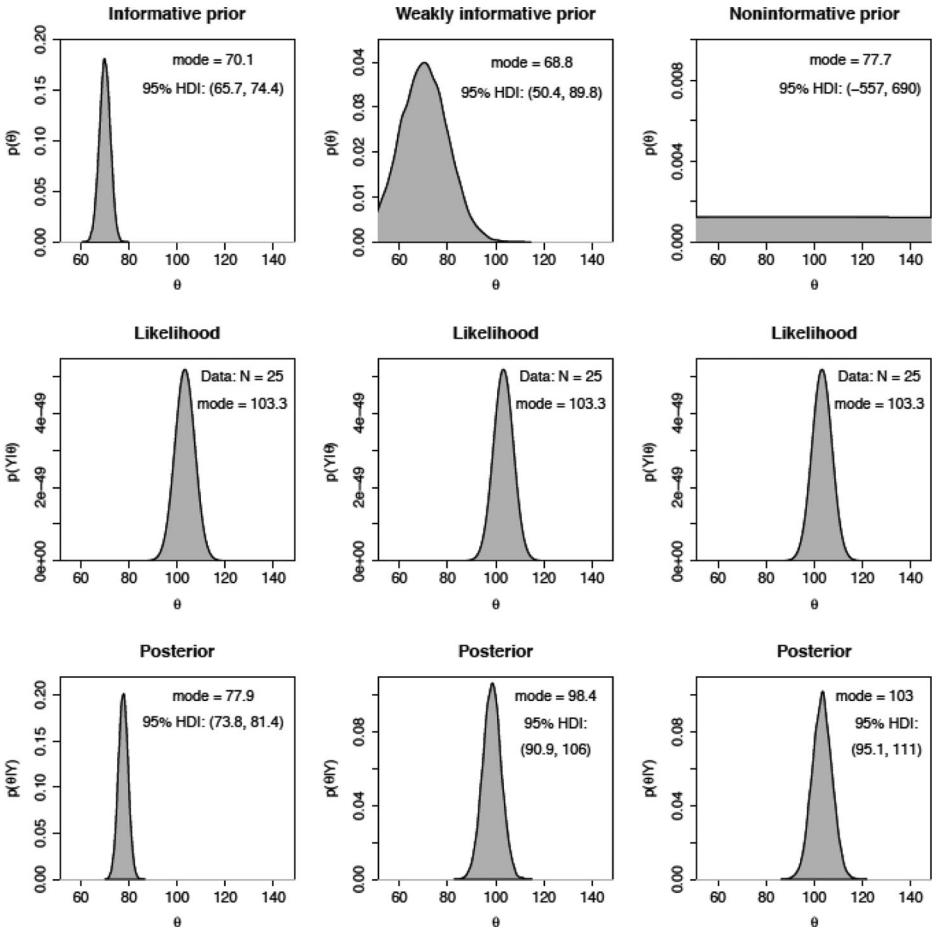
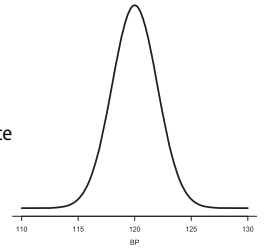


Figure 1. Illustration of how priors with different levels of informativeness impact final model results by influencing the posterior distribution. The plots can be interpreted as follows. Each column maps onto the Bayes' rule equation provided in EQ1. There is a prior, data likelihood, and the resulting posterior. Column 1 shows the impact of an informative prior (with relatively little uncertainty) on the posterior (notice the posterior also has relatively little uncertainty – i.e., the variance is relatively smaller in this posterior). Column 2 illustrates a weakly informative prior with a larger degree of uncertainty compared to the informed prior in Column 1, and Column 3 illustrates a non-informative prior that is quite flat. The non-informative prior has the least amount of impact on the location of the posterior (the mode in the likelihood and the mode in the posterior are both 103 in Column 3). For more information on the impact of priors in latent growth models, see: Depaoli (2012, 2013). Plots for the prior and posterior were created using the rjags package in R (Plummer, 2015).

In contrast, a non-informative prior would capture a much greater deal of uncertainty surrounding the possible population parameter values. An example of this larger degree of uncertainty is depicted in the third column of Figure 1. Non-informative priors can appear rather flat when plotted because they represent a large degree of uncertainty surrounding the possible parameter values. Priors with a larger degree of uncertainty typically have minimal impact on the location of the posterior (i.e., the final answer).

The main question for many applied researchers would likely surround how to *implement* priors within the model estimation code. Of course, each statistical programme uses slightly different code, but the notion for including priors is very similar across programmes. In our *t*-test example looking at systolic blood pressure, we might assume that one group has a mean blood pressure value of 120 ($\mu_1 = 120$, where μ_1 denotes the mean blood pressure for group 1). We can form a prior reflecting this assumption and specify it directly in the code as follows; we also provide another thorough example for conducting a Bayesian analysis in the application section below.

Parameter	Prior code (can be specified in this way in the software programme)	Plot of prior distribution
μ_1 Represents the parameter for the mean systolic blood pressure for group 1	$\mu_1 \sim N(120, 4)$ This code is used in the software syntax to indicate that the prior belief is that the mean blood pressure for group 1 is distributed ($\sim$) normally (N), with a mean of 120 and a variance of 4. This prior can come from previous literature, expert opinions, etc.	 This is a plot of the prior, showing it centres over 120 and has a variance of 4.

The nature of estimation in Bayesian statistics

Unlike frequentist methods, Bayesian methods do not obtain a single point estimate for a population parameter. Instead, a summary of the posterior is used to capture the estimation of a given population parameter. For example, a population parameter could be described by the mean and variance of the posterior distribution (which represents the parameter estimates in

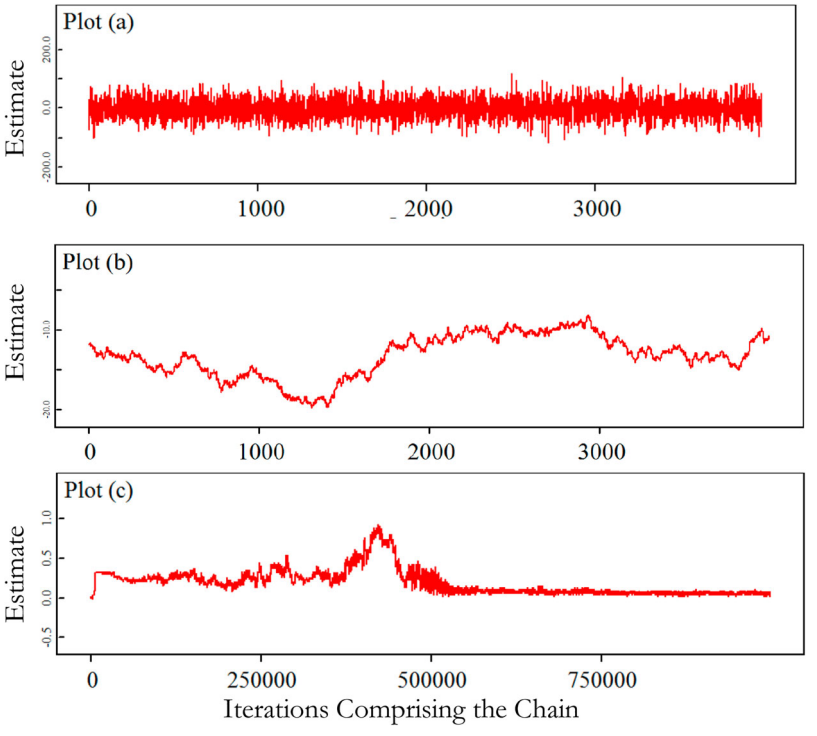


Figure 2. Three convergence plots illustrating evidence of convergence (a), evidence of high autocorrelation with a small effective sample size due to the higher autocorrelation (Kruschke, 2015) (b), and non-convergence becoming stable and converging after 500,000 iterations (c). Each chain is comprised of many samples pulled from the posterior distribution. Once enough samples are extracted, then the posterior distribution can be reconstructed. The mean of these samples then represents the final estimate for the given model parameter being estimated, and the variance of these samples represents the amount of uncertainty surrounding the estimate of the parameter value. Chain information was extracted using the MplusAutomation package (Hallquist & Wiley, 2014) in conjunction with the ggmcmc package (Fernández-i-Marín, 2016) in R.

the model), where the posterior now captures the uncertainty of the population value in a distributional form rather than a point estimate. This result is a key difference in Bayesian and frequentist findings; population parameters in the Bayesian framework are estimated using distributions and not point estimates.

The posterior distribution is not something that can typically be derived in a closed form solution. Whenever the posterior cannot be directly solved for, another process can be used that draws samples from the posterior. In other words, the posterior is reconstructed based on a compilation of many samples that are 'pulled' from the posterior. Each sample pulled is viewed as a likely value that the parameter value can take on. Once one sample is drawn, then another sample is drawn, and so forth. When enough samples have been obtained, the posterior can be accurately approximated through these samples. Then the posterior is converged upon.

The process used here involves generating many representative values (or samples of population values) of the posterior through a Monte Carlo (simulation) process in an attempt to reconstruct the posterior. This feature of deriving the posterior happens though an iterative process and differs substantially from most frequentist estimation methods. In particular, the iterative process for computing the posterior distribution is typically done by implementing Markov chain Monte Carlo (MCMC) techniques. In this process, a Markov chain (i.e., a series of samples drawn from the posterior that are linked to form a 'chain') is constructed for each model parameter using Monte Carlo (simulation) procedures. For example, see Figure 2(a), which presents a chain for a single parameter. Each dot is connected to form a line (or chain). In this plot, the dots represent samples pulled from the

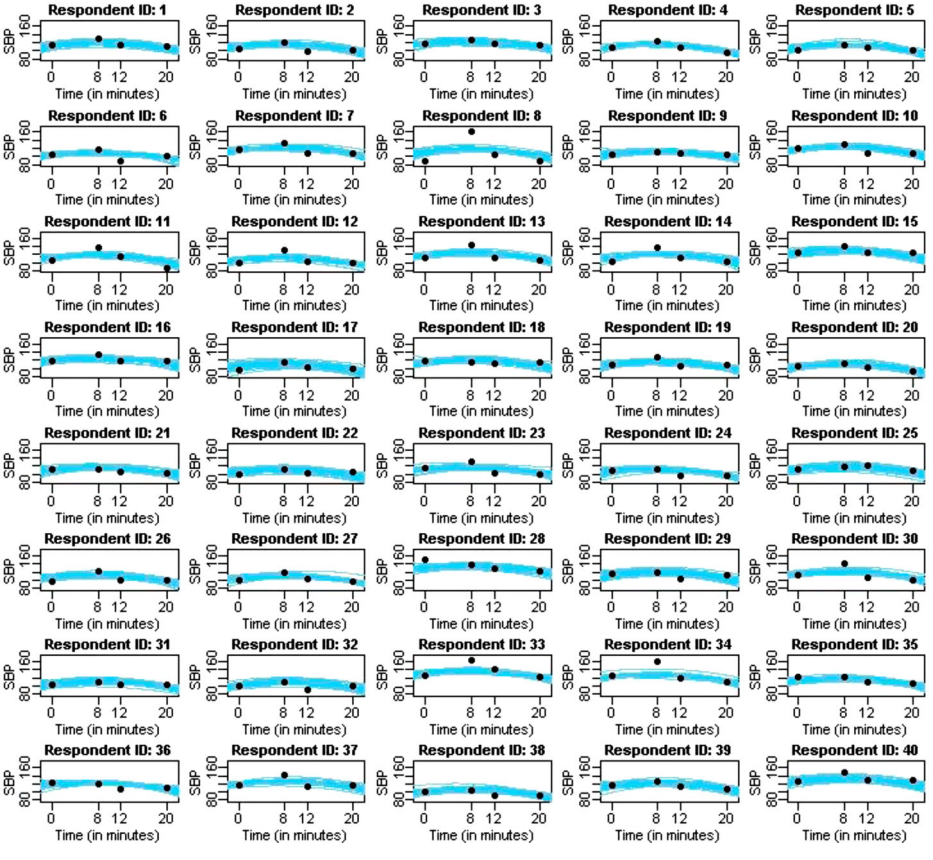


Figure 3. A matrix plot showing the regression lines for each participant, with credible quadratic trends superimposed. This matrix shows that systolic blood pressure follows a quadratic curve across time for the majority of participants. The stressor onset was between 0 and 8 minutes. This plot was created by adapting code from Kruschke (2015).

posterior, where each sample is viewed as a likely value that the parameter can take on. Once enough samples are drawn from the posterior and a stable chain is obtained, then an accurate approximation of the posterior is formed. The mean according to the *y*-axis can be viewed as the mean of the posterior, and the height of the chain represents the amount of variance in the posterior distribution.

Once stability has been established in the chain (i.e., the mean – or horizontal centre – and variance – or height – of the chain are stable, see Figure 2(a)), then the final model estimates are interpreted. Figure 2 shows examples of chains that have converged and those that have not. The beginning portion of the chain is often discarded as the 'unstable' portion (this is called the *burn-in phase*), and the last portion of the chain is used to construct the single correct posterior produced from Bayes' rule via MCMC, where final estimates for model parameters are derived. Characteristics such as the mean, mode, and variance of the sampled posterior distribution are then used to describe the population parameter.

For more details on all of these topics, see the online supplementary material. Next, we provide an example of using Bayesian methods in health psychology in order to illustrate how a researcher can conduct and interpret Bayesian techniques.

Introduction to the acute stress example

An acute stress response and subsequent recovery are commonly studied in a laboratory setting. The stress response can easily be measured with objective physiological markers including blood

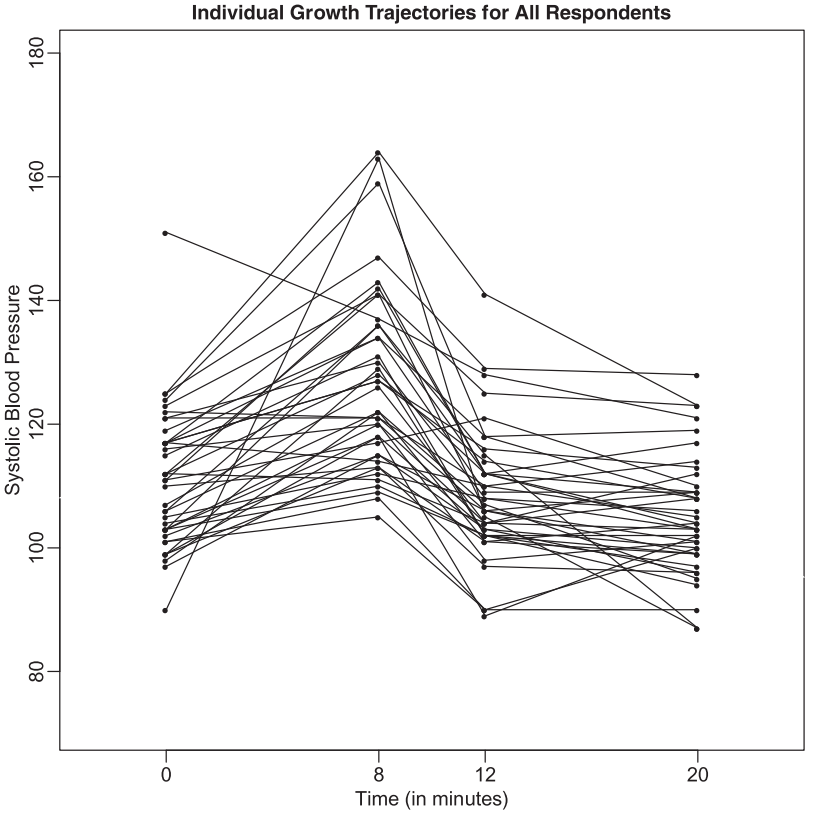


Figure 4. Raw plots of growth trajectories from the 40 participants in the example provided. Upon estimating the latent growth curve model, a single growth trajectory is obtained that conveys the overall pattern of change in systolic blood pressure across time. The stressor onset was between 0 and 8 minutes. This plot was created after extracting chain information using the MplusAutomation package (Hallquist & Wiley, 2014) in conjunction with the ggmcmc package (Fernández-i-Marín, 2016) in R.

pressure. Data used for the present article represent systolic blood pressure (mm/Hg) collected at baseline and at three time-points during recovery after experiencing an acute stressor in a laboratory setting. Measurements were taken with a digital blood pressure monitor at baseline and three post-stressor time-points.

Our sample of $n = 40$ consists of Hispanic students, who completed measures of systolic blood pressure (see Figure 3 for data structure). Hispanics represent the largest and fastest growing minority population in the United States (Krogstad & Lopez, 2014). Despite this, they remain underrepresented in social science research (Knight, Roosa, & Umaña-Taylor, 2009), and particularly in health psychology research, often due to limited access (Yancey, Ortega, & Kumanyika, 2006). However, Hispanics have shown to differ on measures of blood pressure compared to other racial/ethnic groups (Wright, Hughes, Ostchega, Yoon, & Nwankwo, 2011), suggesting they may represent a substantively different group within the population. Only small samples were possible here because data were very expensive to collect. Bayesian methods are used to obtain posterior distributions, which do not require large sample sizes to approximate large sample (frequentist) *p*-values or confidence intervals.

The statistical model implemented

The statistical model used for this inquiry was a latent growth curve model (LGCM) allowing for quadratic change over time across four time-points. In this type of model, individual growth trajectories (Figure 4) are captured by a single growth trajectory with an intercept (starting point) and slope, indicating overall rate of change across participants (all model parameters are at the group-level). A picture of the model can be found in Figure 5. In a traditional LGCM, the observed

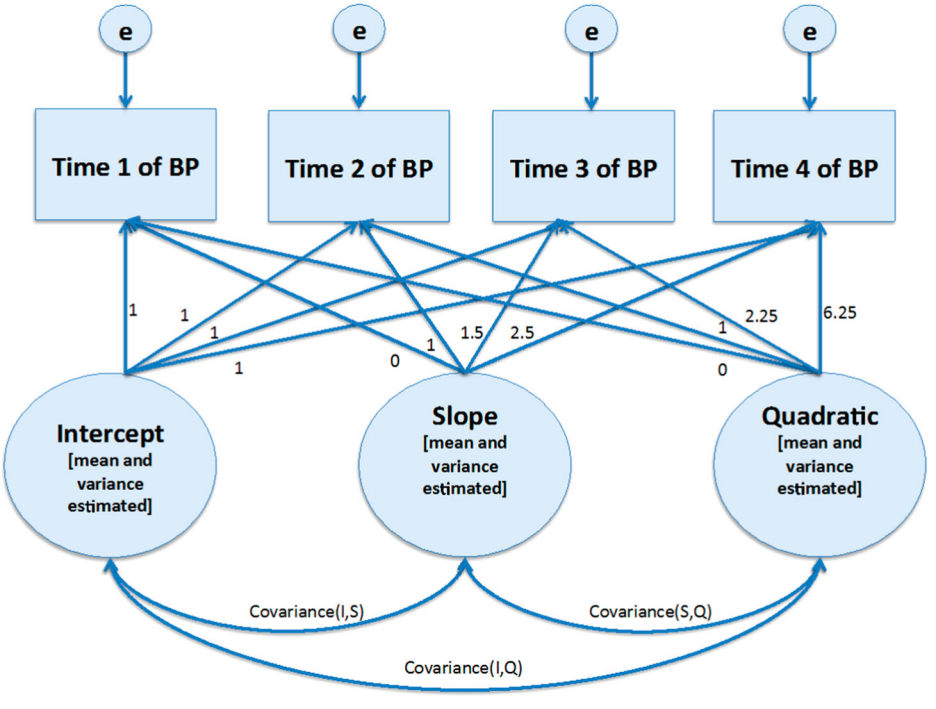


Figure 5. Illustration of the statistical model estimated here: the latent growth curve model with quadratic growth. The three latent growth parameters are the: intercept, slope, and quadratic terms. There are four observed time-points of systolic blood pressure (BP) in the model, each with freely estimated residual variances (e). Finally, covariances were estimated for all growth parameters. The paths have fixed numeric values corresponding with the latent growth curve model with unequal spacing between time-points. See table in online material for more details about all of the parts of this model.

data are represented as squares and the latent variables (called *growth factors*, which correspond to the overall estimated growth trajectory) are represented as circles. In this case there are four measures of blood pressure (BP), each representing a different time-point in the study.[4] It is important to recognise that this is a hierarchal (or nested) model in that time is nested within participants (i.e., each person has four time-points worth of data). There are also three growth factors: intercept, slope, and quadratic terms, which will form the overall growth trajectory across all individuals.[5] We have included a detailed table (online material), which describes all of the model parameters and how they are interpreted. Estimation was conducted via the M*plus* software program version 7.31 (Muthén & Muthén, 1998–2015).

It is important to note that this example is pedagogical in nature and should not be used to draw substantive conclusions. We implemented a basic quadratic LGCM without going through an extensive model building and selection process, as would be done in a substantive inquiry; for a discussion of what to consider when model building for LGCMs, see Curran, Obeidat, and Losardo (2010). We also highlight the ability for the LGCM to reflect unequal time spacing between data collection phases. For the purposes of this example, we coded time-points as baseline, 8 minutes, 12 minutes, and 20 minutes post-baseline, with the stressor onset between 0 and 8 minutes. Results can be used as a guide for implementing and interpreting Bayesian findings, which can be generalised to other types of models.

Table 1. Annotated syntax for conducting a Bayesian LGCM: an example using systolic blood pressure over time.

Code in *Mplus* for the LGCM	Description of code
title: Hispanic systolic BP Bayesian LGCM	Create a title for the analysis being conducted
data: file is data.txt	Indicate the data file being used; data.txt is the name of this data file
variable: names are bp1-bp4;	Name the four time-points of data; we named them 'bp' here to represent the four measures of systolic blood pressure
analysis:	Setting up the details of estimation must start with the line 'analysis'
estimator = BAYES;	Request the Bayesian estimator
fbiterations = 40000;	We requested 40K iterations in the chain, the first half are discarded as burn-in, and the second half used as the final posterior where the final estimate is derived
model:	Specifying the latent growth curve model starts with the line 'model'
y1-y4*;	Estimating residual variances for each time-point (estimation is denoted by *)
i s q | bp1@0 bp2@1 bp3@1.5 bp4@2.5;	Setting up the growth model. It is a quadratic model with an intercept (i), slope (s), and quadratic term (q). The four data time-points are used (bp1–bp4), and the numbers on the right side of the @ indicate unequal time spacing between the data points (i.e., 0, 1, 1.5, and 2.5 – this spacing reflects that time-points bp2 and bp3 were collected closer together than the other time-points)
i*; s*; q*;	Estimating variances for growth parameters
[i*](a1);	Estimating intercept mean (means are denoted by including the code in a bracket, '[i*]'); the prior is indicated with 'a1' as the identification name
[s*](b1);	Slope mean; the prior is indicated with 'b1' as the identification name
[q*](c1);	Quadratic mean; the prior is indicated with 'c1' as the identification name
model priors:	Specifying priors in the model
a1∼N(115,10);	Prior for the intercept mean, using the 'a1' identification name to denote the intercept. It is a normal prior (denoted by 'N'), with a mean of 115 and a variance of 10. These numbers were derived from the literature (see also Table 2)
b1∼N(23.29,10);	Prior for the slope mean, using the 'b1' identification name to denote the slope. It is a normal prior with a mean of 23.29 and a variance of 10. These numbers were derived from the literature (see also Table 2)
c1∼N(-6.94,100);	Prior for the quadratic mean, using the 'c1' identification name to denote the quadratic term. It is a normal prior with a mean of −6.94 and a variance of 100. These numbers were derived from the literature (see also Table 2)
plot: type = plot2;	Requesting plots (e.g., to show chain convergence)

Note: Bolded text denotes the code needed to convert the base, traditional LCGM code into being Bayesian. Notice that there are only two sections that need additional code to make this a Bayesian model: (1) under the 'analysis' section and (2) the 'model priors' section.

Table 2. Model priors implemented in the current study.

Parameters	Distributional form of the priors and values of parameters in priors	Type of prior (non-, weakly, highly informative)	Source of background information and justification of use	Potential impact on final outcome
Intercept Mean	Normal Distribution Mean = 115 Variance = 10 Prior: *N*(115,10)	Weakly Informative	CDC report (Wright et al., 2011) on mean systolic blood pressure in Mexican-American adults age 18–39. The CDC is viewed as a leading source for guidelines on systolic blood pressure and is one of the few sources with blood pressure data for Mexican-Americans, and we pulled the mean of the prior (115) from this source. The variance for the priors for the intercept, slope, and quadratic means were selected in conjunction with an expert (i.e., the last author of this article) on systolic blood pressure and the impact of the acute stressor implemented here	We pulled information from the CDC on expected (mean) systolic blood pressure for Mexican-Americans. With a variance of the prior set to 10, the prior was thought to be weak enough to allow the data to 'speak' and not predetermine the final outcome for the Intercept mean being estimated. To assess full impact of the prior, we also conducted a sensitivity analysis to ensure that the impact of the prior was clearly understood. The variance value for the prior was specifically chosen by and expert to be weakly informative as to ensure that the prior would not have an unintended impact on the posterior
Slope Mean	Normal Distribution Mean = 23.29 Variance = 10 Prior: *N*(23.29, 10)	Weakly Informative	Acute social stress paradigm measuring blood pressure at 10, 20, and 40 minutes post-stressor onset (Izawa et al., 2008). Izawa et al. used the same physiological markers and the same stressor as the current investigation. Given that blood pressure has a predictable response to acute stress, and this change does not differ across ethnicities, the Izawa et al. study was thought to provide reasonable priors for linear change in blood pressure. Thus, we pulled the mean for the prior on the slope from this study. The variance of the prior was derived from an expert	We pulled the mean of the prior from Izawa et al. (2008). The variance value for the prior was specifically chosen by an expert to be weakly informative as to ensure that the prior would not have an unintended impact on the posterior. In addition, we preformed a prior sensitivity analysis for this parameter to ensure that impact of the prior was clearly understood and that the data were having influence on final model outcomes
Quadratic Mean	Normal Distribution Mean = −6.94 Variance = 100 *N*(−6.94, 100)	Relatively non-informative	Akin to the slope term, the mean for this prior was derived from Izawa et al. (2008). It was thought that reasonable information for Hispanic participants could be derived from the study on Japanese participants who experienced the TSST. The variance of the prior was derived from an expert	We pulled the mean of the prior from Izawa et al. (2008). The variance value for the prior was specifically chosen by an expert to be weakly informative as to ensure that the prior would not have an unintended impact on the posterior. In addition, we preformed a prior sensitivity analysis for this parameter to ensure that impact of the prior was clearly understood and that the data were having influence on final model outcomes
Residual variances	Inverse gamma distribution IG(−1,0)	Non-informative	We had no prior information about the residual variances for outcomes so a non-informative prior was implemented; see Muthén and Muthén (1998–2015) for more details on this prior	Given that this prior is specified as non-informative, it was thought to have little impact on the final model estimates
Growth parameter variance/ covariance matrix	Inverse Wishart distribution IW(0,−4)	Non-informative	We had no prior information about the growth parameter variances and covariances so a non-informative prior was implemented; see Muthén and Muthén (1998–2015) for more details on this prior	Given that this prior is specified as non-informative, it was thought to have little impact on the final model estimates

Note: CDC, Centers for Disease Control; TSST, Trier Social Stress Test. For more information on how priors were derived, please contact the first author. We also note that there are a variety of methods that can be used for deriving priors from previous research or other sources. For example, priors can be derived from: (1) using previous research findings (e.g., Gelman, Bois, & Jiang, 1996) or a meta-analysis, (2) data-splitting techniques, (3) data-driven priors (e.g., Darnieder, 2011; Richardson & Green, 1997; Wasserman, 2000), and (4) expert elicitation (e.g., O'Hagan et al., 2006). As long as the prior can be substantively or theoretically justified, then most origins of that prior are appropriate. The methods used for deriving priors in the current investigation are only a single example of the possible methods that could have been implemented. See previous citations for details on other methods.

How to implement this model through a Bayesian perspective

There are a variety of different software programmes that can be used to estimate models through the Bayesian framework (e.g., the R programming environment, OpenBUGS, and M*plus*). Although they all have their own unique syntax for conducting the analysis, there are similarities across all of them with respect to how priors are incorporated into the model. In order to provide a concrete example for how to conduct a Bayesian analysis, we have included annotated code for the current example in Table 1. We used the M*plus* software program because it is relatively straightforward for conducting Bayesian analysis and requires only a few extra lines of code. In Table 1, we have included all code needed to estimate a Bayesian LGCM. The lines of code in regular (i.e., not bolded) text represent the base code needed to conduct the LGCM in the frequentist (conventional) estimation framework. All bolded lines of coded are required to make this model Bayesian. Notice that the extra lines of code to convert this to a Bayesian model are minimal. The main addition is the specification of priors. In order to specify a prior on a model parameter (e.g., the intercept of the growth trajectory), the model parameter must be 'named' in the code. In our example, we have created an identification name of 'a1' to represent the intercept; this is denoted as [i*](a1), where '(a1)' is the identification name linked to the intercept which is coded as '[i*]' in M*plus*. In the model priors command, that identification name is used to specify the prior, in this case 'a1∼*N* (115, 10)' – indicating a1 (i.e., the intercept) is distributed (∼) as a normal distribution (*N*) that is centred at 115 with a variance of 10. This is the main addition to converting the base code to be Bayesian.

We recognise that not all readers will necessarily be familiar with this model or with this software programme. As a result, we have included a simple Bayesian example in the online material using a regression model in nine different Bayesian programmes so readers can reference software they are

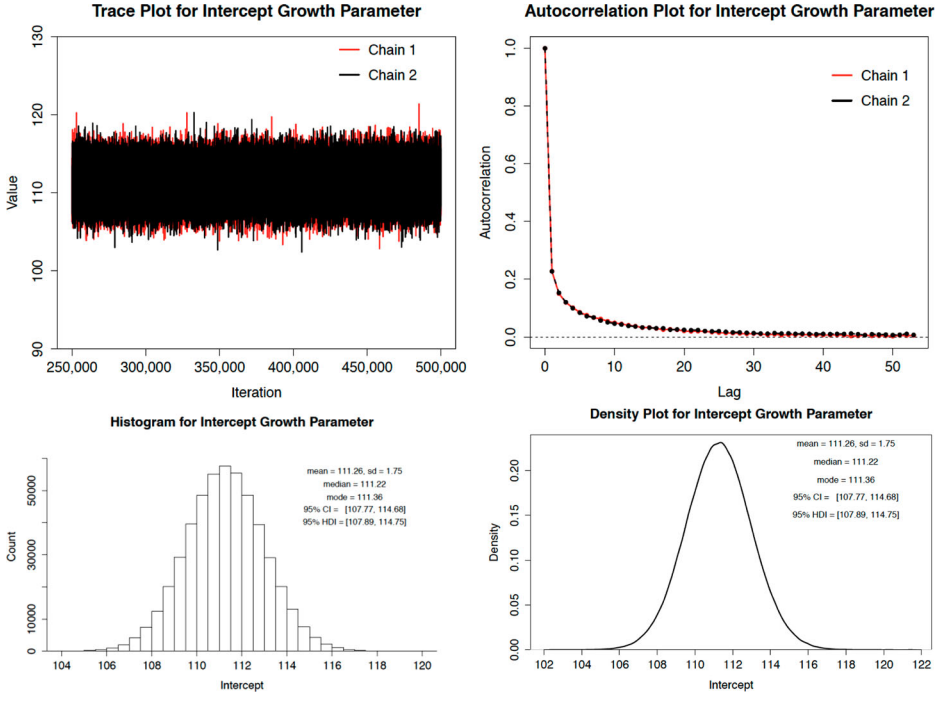


Figure 6. Convergence, posterior histogram, posterior kernel density, and autocorrelation plots (two chains) for the Intercept mean parameter with 500,000 chain iterations. All other parameters showed a similar pattern of results and plots for them can be found in the online supplementary material. These plots were created after extracting chain information using the MplusAutomation package (Hallquist & Wiley, 2014) and ggmcmc packages (Fernández-i-Marín, 2016) in R.

most familiar with.[6] Full code and model results for the current example in *Mplus* are also provided in this online material.

Results

Before actually analysing the LCGM, all priors must be thoroughly defined and we include this information in Table 2. Specifically, priors for the growth parameters were derived from multiple sources, including the CDC and similar stress response studies.

Two MCMC chains were specified for each model parameter, each with disparate starting values. Chain convergence was obtained with 500,000 total iterations (first half removed as burn-in).[7] We examined chain convergence using the Gelman and Rubin (1992) convergence diagnostic called the potential scale reduction factor. Essentially, this diagnostic creates a ratio of between- and within-chain variability to examine whether the two chains have converged together; if the ratio is near 1.0 (e.g., <1.05), then the chains have likely converged with one another. Figure 6 shows the following plots for the intercept mean: convergence, posterior histogram, posterior kernel density, and autocorrelation plots for both chains. The convergence plot shows stability in the chain, and the histogram and density plots show a smoothed, normally distributed posterior. Additionally, the autocorrelation plots illustrate low degrees of dependency within each chain, which is optimal given that high dependency can be a sign of a poorly formed (i.e., *mixed*) chains or even model misspecification. Similar plots for all model parameters are available in the online material.

Table 3. Final model results for the latent growth curve model on systolic blood pressure in Hispanic participants after an acute stressor; based on 500,000 MCMC iterations.

	Posterior median estimate	Posterior mean estimate	Posterior standard deviation	95% Bayesian credible interval (equal tail probabilities)	95% Highest density interval	Effective sample size
Intercept (I)						
Mean	111.26	111.26	1.75	(107.85, 114.72)	(107.88, 114.75)	122,713.46
Variance	129.18	134.78	42.80	(66.87, 234.29)	(58.31, 220.05)	51,724.55
Slope (S)						
Mean	16.12	16.11	3.33	(9.56, 22.58)	(9.58, 22.59)	19,937.05
Variance	116.20	136.63	93.47	(19.23, 372.35)	(5.28, 317.53)	20,986.57
Quadratic (Q)						
Mean	−7.42	−7.42	1.36	(−10.05, −4.75)	(−10.06, −4.76)	19,076.82
Variance	16.93	20.12	14.16	(2.66, 55.98)	(0.64, 47.61)	21,466.99
Covariances						
I with S	−51.38	−55.68	50.97	(−169.05, 33.35)	(−160.57, 39.55)	27,992.07
I with Q	15.42	16.76	19.26	(−17.99, 58.92)	(−19.61, 56.80)	30,357.76
S with Q	−42.26	−51.30	36.05	(−142.40, −6.52)	(−121.04, −1.06)	21,613.03
Residual var.						
Time 1	19.28	22.64	17.06	(0.94, 63.16)	(0.000, 54.91)	11,798.54
Time 2	249.97	262.09	84.44	(132.85, 461.32)	(116.55, 432.53)	45,858.63
Time 3	143.69	153.48	66.63	(52.31, 310.31)	(39.78, 288.06)	27,683.77
Time 4	13.56	16.26	12.73	(0.55, 46.83)	(0.00, 40.53)	11,174.84

Note: Posterior median estimate = the median of the MCMC chain estimated for each model parameter. Posterior mean estimate = the mean of the MCMC chain estimated for each model parameter. Posterior standard deviation = the standard deviation of the posterior distribution that resulted for each model parameter. 95% Credible interval = akin to a frequentist confidence interval, but interpreted in a Bayesian way (see van de Schoot & Depaoli, 2014 for more details). Note that this interval has been computed with equal tail percentages. Another type of interval is called the highest density interval, which we also present in Figure 5 for the intercept, slope, and quadratic means. Note that the highest density intervals were computed using the HDIofMCMC() R function discussed in Kruschke (2015), which is also available in the BEST package in R (Kruschke & Meredith, 2015). Effective sample sizes take into account the amount of autocorrelation in the chain. If autocorrelation is high, then this effective sample size of the chain is going to be lower in order to account for the high degree of dependency among the samples in the chain. Effective sample size values should be at least 10,000 to ensure that credible intervals are stable (Kruschke, 2015, Section 7.5.2, pp. 182+). Growth factor covariances are included to be consistent with conventional estimation of LGCMs through the structural equation modelling framework to improve the accuracy of the estimated latent growth trajectory (Bollen & Curran, 2006, p. 22). However, the online supplementary material includes code where the covariances are restricted to zero to more closely mimic convention in the multilevel modelling framework.

Final model estimates are presented in Table 3. The intercept indicates that Hispanic participants had an average baseline systolic blood pressure of 111.26 (sd = 1.75 for the posterior distribution, which captures the variability in the posterior; see Figure 6 for a picture of the posterior density). The linear slope estimate of 16.12 (sd = 3.33) indicates an increase in blood pressure over the course of experiencing the stressor. The quadratic slope of −7.42 (sd = 1.36) indicates that over

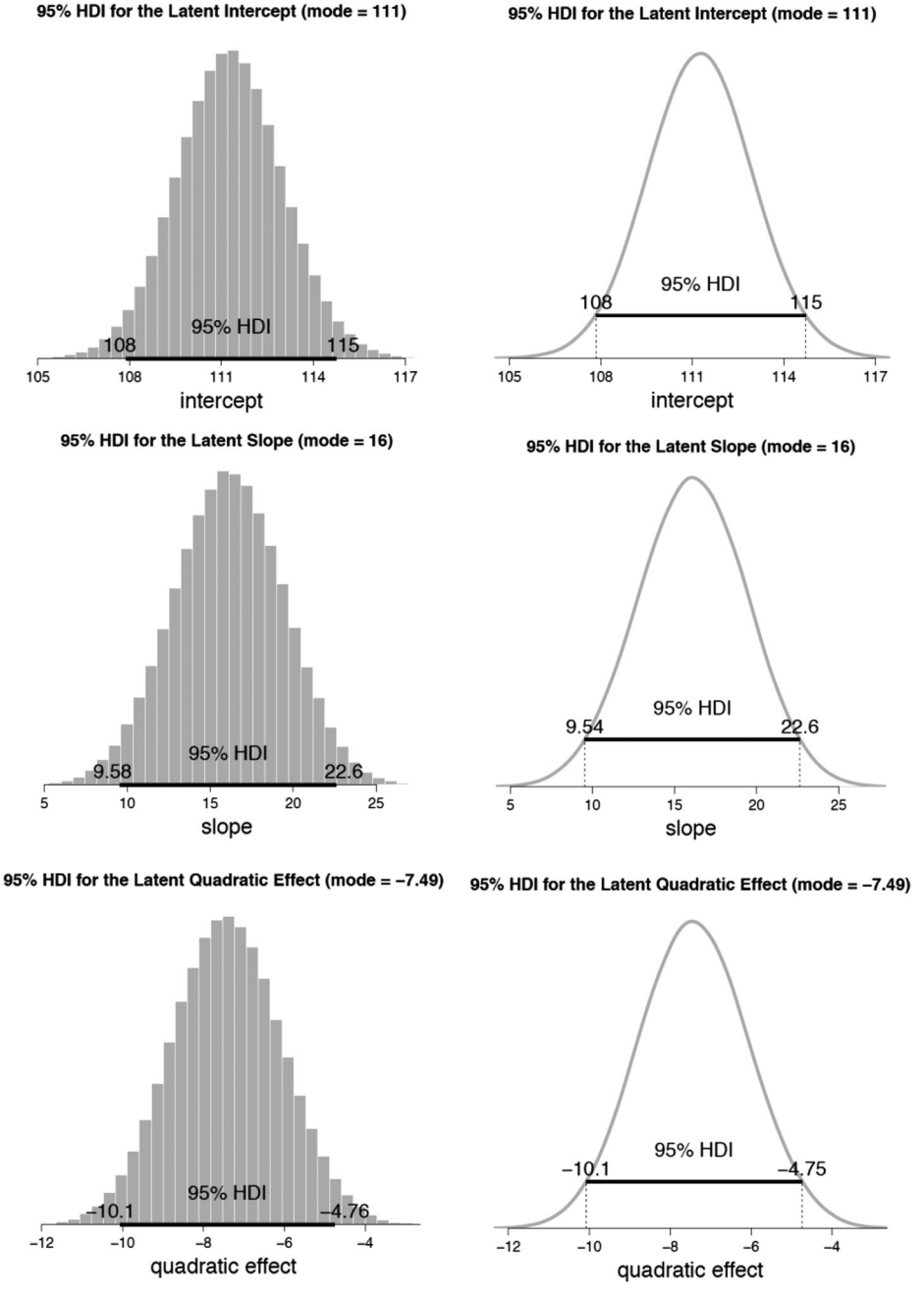


Figure 7. Highest density interval plots in histogram and smoothed forms for the intercept, slope, and quadratic means. These plots incorporate the mode of the posterior, but plots based on the mean or median would look comparable due to the symmetry of the posteriors. The central tendency of the posterior can be visualised, as well as the amount of uncertainty surrounding this estimate (i.e., the width of the distribution). These plots were created using the BEST package in R (Kruschke & Meredith, 2015).

time this blood pressure level dropped, which would be expected when recovering from a stressor. In order to capture the richness of Bayesian results, we have included highest density interval (HDI) plots in Figure 7, which can also be used to depict final model estimates. HDI plots are presented in histogram and smoothed forms for the intercept, slope, and quadratic means. These plots show the richness of Bayesian results in that they illustrate areas under the posterior distribution that are more likely for each of the model parameters; this interpretation is in contrast to point estimates obtained with frequentist approaches. An interval was specified using the 95% credible interval surrounding the mode of the posterior, and these intervals provide an indication of values that are more plausible for the parameter value compared to values outside of the interval. For example, the HDI for the intercept mean shows that values outside of the systolic blood pressure range of 108–115 would be considered implausible. Finally, we revisit Figure 3, which illustrates this estimated growth trajectory for participants in the study, with 95% confidence bands added to show the uncertainty associated with the average trajectory. Likewise, Figure 3 shows plots for all participants, with credible quadratic trends superimposed. Overall, there is a quadratic decrease in blood pressure post-stressor for most participants, which can be further illustrated in the estimated growth trajectory in Figure 8. Based on patterns in Figures 4 and 8, it is clear that some participants did not respond to the stressor with elevated systolic blood pressure. Specifically, some participants showed the expected increase at time-point 2, with a decrease at time-points 3 and 4. Whereas, other participants did not show elevations in their systolic blood pressure, seeming to 'flat-line' in terms of systolic blood pressure. This is not an uncommon response to an acute social stressor, and could be explained by, for example, perceived social support or social status; such moderators can be added in subsequent analyses in the Bayesian framework.

A full sensitivity analysis of priors is presented in the online supplementary material; space limitations prevent it from appearing here, but a sensitivity analysis should always be conducted with informed priors. Depaoli and van de Schoot (2017) describe in more detail what to report in a Bayesian analysis.

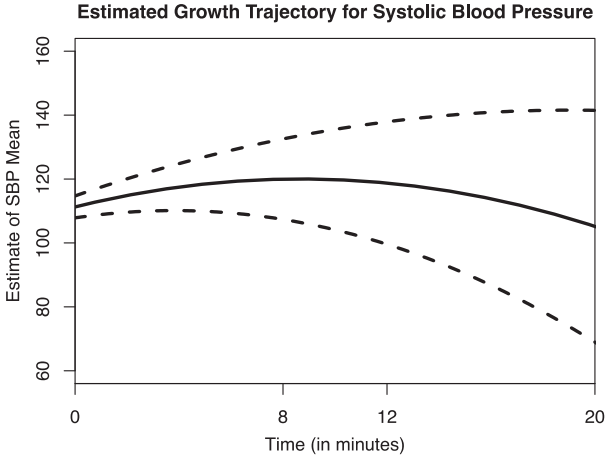


Figure 8. The estimated growth trajectory for systolic blood pressure changes upon experiencing an acute stressor in Hispanic participants after baseline (Time = 0). Time spacing was treated unequal at 8-minutes, 12-minutes, and 20-minutes post-baseline. The stressor onset was between 0 and 8 minutes. Notice the dashed lines represent lines based on 95% credible intervals for the intercept, slope, and quadratic terms. The credible intervals were relatively larger for the slope and quadratic terms compared to the intercept due to the larger deviation in the curves along the *x*-axis (i.e., over time, or as time progresses). This plot was created using base code in R.

Conclusion

As the use of Bayesian statistics continues to rise due to the increase in user-friendly software and tutorials, a large burden will be placed on journal and grant reviewers to ensure proper implementation of Bayesian methods. To aid in this matter, we have provided a comprehensive checklist that reviewers can use when evaluating a manuscript or grant proposal; this checklist is presented in Table 4. Likewise, we include several main points that need to be addressed when conducting a Bayesian analysis. The following represent important points to address, as described further in Kruschke (2015):

- provide motivation for why Bayesian methods are being used (e.g., small samples, benefits of using priors),
- clearly describe the model and its parameters (e.g., latent growth model with intercept and slope terms),
- describe and justify the priors (e.g., where informed priors originated from),
- select the software and be aware of default Bayesian settings,
- report the MCMC details (e.g., convergence, number of burn-in iterations in the chain),
- interpret the posterior (e.g., through HDI plots), and
- conduct a sensitivity analysis on informed (subjective) priors.

In our view, health psychology is a perfect setting for implementing Bayesian inference and deriving informed priors. This field is so rich with information and intriguing research that Bayesian methods can really shine through the use of informed priors. Specifically, we believe that incorporating knowledge from previous health-related research can greatly benefit the current work being done in the field. The main point to keep in mind as the field of health psychology begins to utilise these methods more frequently is that transparency and openness are key in Bayesian statistics, especially with respect to the specification and reporting of priors.

Table 4. Checklist reviewers can use when examining a Bayesian manuscript or grant proposal.

What to check for as a reviewer	Present?	Points to consider during evaluation
(1) Is the model clearly defined, with all model parameters identified and explained?	Yes/no	This information must be present so that readers can fully understand the model being estimated
(2) Is there information about the priors used (type of prior, where the prior came from, what the hyperparameter values were)?	Yes/no	If information is not provided about the priors, then there is not enough information to judge the merit of the rest of the article or proposal
(3) Which software and MCMC sampling method was used?	Yes/no	This information should be present so that future readers can reconstruct all analyses
(4) How did the authors assess convergence? Did they do a thorough job to ensure convergence was consistent and stable, even with a longer chain? Further, did they detail the number of settings thoroughly (e.g., number of chains, length of burn-in and post burn-in portions of the chain; see Kaplan & Depaoli, 2013, or Van de Schoot et al. 2014, for an explanation of these settings)?	Yes/no	The method for assessing convergence must be stated. It is important that a thorough assessment of convergence is present. Results will be meaningless without this information. All listed information should be present so that future readers can reconstruct all analyses
(5) Did the authors investigate the influence of the prior specifications on the final model results through a detailed sensitivity analysis? Specifically, if informed priors were used in parameter estimation, then a sensitivity analysis should be reported. Similarly, a sensitivity analysis should be reported if Bayes factors were computed. This step ensures an assessment of the sensitivity of results to the choice of the prior. (Findings should be discussed in the Conclusion section of the article)	Yes/no	If a sensitivity analysis was not conducted on informed priors, then it is not clear how much of an impact the priors may be having on model results. This is an imperative step to ensure conclusions are not being over-stated or exaggerated based on findings

Notes

1. Specifically, maximum likelihood estimates can be obtained with small sample sizes, but the estimates themselves can be biased (i.e., inaccurate). Likewise, *p*-values, standard errors, and the corresponding confidence intervals can be poorly approximated. Bayesian methods have been shown to provide more accurate results in comparison.
2. Readers new to Bayesian terminology may find some of the concepts difficult to distinguish from frequentist concepts. In particular, it is important to note there are substantial differences between frequentist sampling distributions and Bayesian priors. Sampling distributions are theoretical distributions that capture the probability of a statistic, and they are typically used to assess statistical significance in a null hypothesis test. In contrast, priors are used to capture distributions over parameters, where a researcher's beliefs about a population parameter can be summarized by a distribution with a central tendency and a level of uncertainty (e.g., variance) surrounding the likely values of the parameter.
3. For a normally distributed prior, its informativeness can be determined in terms of the mean and the variance of that prior. The mean and the variance of the distribution capture the amount of (un)certainty in the prior (i.e., how (un)certain the researcher is of the likely values of a model parameter). The mean of the prior will provide information about the location of the parameter. The variance of the prior will provide information about the (un)certainty surrounding the location of the parameter. If the researcher is highly certain of the parameter value, then a very informative prior might be specified with a particular location (mean) and a good deal of certainty (i.e., small variance) surrounding that mean. In this case, the mean and the variance contribute to the informativeness of the prior. The prior can still be regarded as informative even if the mean of the prior is inaccurate to the population value – it is still providing information about the location of the parameter, regardless of the accuracy.
4. Data were structured such that these time-points represented four columns of data, and each row was an individual.
5. This type of latent variable model is typically used to find overall (e.g., averaged) growth or change over time for all participants. There are other hierarchical models that can be used to examine individual growth trajectories, but these models are beyond the scope of the current article. Instead, we focus on Bayesian interpretation of a model parameter, such as the intercept, of this growth model.
6. If a reader is less familiar with the LGCM, then just focus here on the 'Bayesian way' of interpreting model results.
7. Note that chain convergence was obtained according to the Gelman and Rubin potential scale reduction factor (PSRF) at 40,000 total iterations in the chain. Depending on the software being implemented, convergence is typically checked after a set number of iterations (e.g., 40,000) rather than determined at the point of a criterion being met for a convergence diagnostic (e.g., the PSRF). However, we increased the number of iterations substantially beyond this initial value of 40,000 (increased to 500,000 iterations) in order to increase the effective sample size (ESS) values for each chain (seen in Table 3) to exceed 10,000; this ensured that estimates for the credible intervals are stable. Convergence remained with this substantial increase in the length of the chain.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J., Fiedler, K., … Nosek, B. A. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, *27*, 108–119.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation approach*. Hoboken, NJ: Wiley.

Bolstad, W. M. (2007). *Introduction to Bayesian statistics*. Hoboken, NJ: Wiley.

Carlin, B. P., & Louis, T. A. (2009). *Bayesian methods for data analysis*. Boca Raton, FL: Chapman & Hall/CRC.

Curran, P., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development*, *11*, 121–136.

Darnieder, W. F. (2011). *Bayesian methods for data-dependent priors* (Dissertation). Ohio State University.

Depaoli, S. (2012). Measurement and structural model class separation in mixture CFA: ML/EM versus MCMC. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*, 178–203.

Depaoli, S. (2013). Mixture class recovery in GMM under varying degrees of class separation: Frequentist versus Bayesian estimation. *Psychological Methods*, *18*(2), 186–219.

Depaoli, S., & Clifton, J. (2015). A Bayesian approach to multilevel structural equation modeling with continuous and dichotomous outcomes. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*, 327–351.

Depaoli, S., & van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics. The WAMBS-checklist. *Psychological Methods, 22*, 240–261.

Fernández-i-Marín, X. (2016). ggmcmc [v0.7.3]: Tools for analyzing MCMC simulations from Bayesian inference. Retrieved from http://xavier-fim.net/packages/ggmcmc

Gelman, A., Bois, F., & Jiang, J. (1996). Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *Journal of the American Statistical Association*, *91*, 1400–1412.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. New York, NY: Chapman & Hall/CRC.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–472.

Hallquist, M., & Wiley, J. (2014). MplusAutomation: Automating Mplus model estimation and interpretation. R package version 0.6-3. Retrieved from https://CRAN.R-project.org/package=MplusAutomation

Hoff, P. D. (2009). *A first course in Bayesian statistical methods*. New York, NY: Springer.

Hox, J., van de Schoot, R., & Matthijsse, S. (2012). How few countries will do? Comparative survey from a Bayesian perspective. *Survey Research Methods*, *6*, 87–93.

Izawa, S., Sugaya, N., Shirotsuki, K., Yamada, K. C., Ogawa, N., Ouchi, Y., … Nomura, S. (2008). Salivary dehydroepiandrosterone secretion in response to acute psychosocial stress and its correlations with biological and psychological changes. *Biological Psychology*, *79*, 294–298.

Kaplan, D., & Depaoli, S. (2013). Bayesian statistical methods. In T. D. Little (Ed.), *Oxford handbook of quantitative methods* (pp. 407–437). Oxford: Oxford University Press.

Kim, S. Y., Suh, Y., Kim, J. S., Albanese, M., & Langer, M. M. (2013). Single and multiple ability estimation in the SEM framework: A non-informative Bayesian estimation approach. *Multivariate Behavioral Research*, *48*, 563–591.

Knight, G. P., Roosa, M. W., & Umaña-Taylor, A. J. (2009). *Studying ethnic minority and economically disadvantaged populations: Methodological challenges and best practices*. Washington, DC: American Psychological Association.

Krogstad, J. M., & Lopez, M. H. (2014, April). Hispanic nativity shift. Pew Research Center. Retrieved from http://www.pewhispanic.org/2014/04/29/hispanic-nativity-shift/

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test.. *Journal of Experimental Psychology: General*, *142*, 573–603.

Kruschke, J. K. (2015). *Doing Bayesian analysis: A tutorial with R, Jags, and STAN*. San Diego, CA: Academic Press.

Kruschke, J. K., & Meredith, M. (2015). Package 'BEST'. Retrieved from https://cran.r-project.org/web/packages/BEST/BEST.pdf

Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R., & Jones, D. R. (2005). How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine*, *24*(15), 2401–2428.

Muthén, L. K., & Muthén, B. (1998–2015). *Mplus user's guide* (7th edition). Los Angeles, CA: Muthén & Muthén.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., & Rakow, T. (2006). *Uncertain judgments: Eliciting experts' probabilities*. West Sussex: Wiley.

Plummer, M. (2015). rjags: Bayesian graphical models using MCMC. R package version 3-15. Retrieved from https://CRAN.R-project.org/package=rjags

Richardson, S., & Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *59*, 731–792.

van de Schoot, R., Broere, J. J., Perryck, K. H., Zondervan-Zwijnenburg, M., & van Loey, N. E. (2015). Analyzing small data sets using Bayesian estimation: The case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *European Journal of Psychotraumatology*, *6*(0), 25216.

van de Schoot, R., & Depaoli, S. (2014). Bayesian analyses: Where to start and what to report. *European Health Psychology*, *16*, 75–84.

van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*, *85*, 842–860.

van de Schoot, R., Winter, S., Zondervan-Zwijnenburg, M., Ryan, O., & Depaoli, S. (2017). A systematic review of Bayesian papers in psychology: The last 25 years. *Psychological Methods, 22*, 217–239.

Wasserman, L. (2000). Asymptotic inference for mixture models by using data-dependent priors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *62*, 159–180.

Wright, J. D., Hughes, J. P., Ostchega, Y., Yoon, S. S., & Nwankwo, T. (2011). *Mean systolic and diastolic blood pressure in adults aged 18 and over in the United States, 2001–2008*. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

Yancey, A. K., Ortega, A. N., & Kumanyika, S. K. (2006). Effective recruitment and retention of minority research participants. *Annual Review of Public Health*, *27*, 1–28.

Zhang, Z., Hamagami, F., Wang, L., Grimm, K. J., & Nesselroade, J. R. (2007). Bayesian analysis of longitudinal data using growth curve models. *International Journal of Behavioral Development*, *31*, 374–383.