Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 29 Number 1, January 2024 ISSN 1531-7714

Clearer Analysis, Interpretation, and Communication in Organizational Research: A Bayesian Guide

Karyssa A. Courey, *Rice University*, Frederick L. Oswald, *Rice University*, and Steven A. Culpepper, *University of Illinois Urbana-Champaign*

Historically, organizational researchers have fully embraced frequentist statistics and null hypothesis significance testing (NHST). Bayesian statistics is an underused alternative paradigm offering numerous benefits for organizational researchers and practitioners: e.g., accumulating direct evidence for the null hypothesis (vs. 'fail to reject the null'), capturing uncertainty across a distribution of population parameters (vs. a 95% confidence interval on a single point estimate) – and through these benefits, communicating statistical findings more clearly. Although organizational methodologists in the past have promoted Bayesian methods, only now is easy-to-use JASP statistical software available for more widespread implementation. Moreover, the software is free to download and use, is menu-driven, and is supported by an active multidisciplinary user community. Using JASP, our tutorial compares and contrasts frequentist and Bayesian approaches for two analyses: a multiple linear regression analysis and a linear mixed regression analysis.

Keywords: Bayesian statistics, Bayesian regression, Bayesian linear mixed model, statistical communication

Introduction

Statistics can often be frustrating for people before they become illuminating. First-year organizational science graduate students and organizational researchers are often taught null hypothesis significance testing (NHST) in the frequentist paradigm as a primary tool for conducting and publishing their research. For instance, under NHST, we learn how we cannot "accept the null" under nonsignificant findings; instead, we must simply "fail to reject the null." What if there were a different way of conducting statistics, where one can:

- provide the most probable parameter estimates for a variable in a given model (versus a single observed point estimates and its 95% confidence interval [CI]);
- accumulate evidence in direct support of the null hypothesis (versus indirectly by "failing to reject the null" in NHST);
- accumulate evidence to compare any two models against one another;
- communicate the probability of obtaining statistical findings in a simple and straightforward manner (versus relying on awkward interpretations of *p* values and 95% CI);

Indeed, we can do all of this and more with Bayesian analyses. The current paper focuses specifically on Bayesian methods because although they have been introduced and demonstrated in previously published organizational science papers (e.g., Jebb & Woo, 2015; Kruschke et al., 2012; Zyphur & Oswald, 2015), they are still rarely adopted, compared to other fields of research, such as in the physical sciences and life sciences (van de Schoot et al., 2017). However, an increased appreciation for Bayesian analyses in organizational research is a promising development, coupled with the availability of conducting Bayesian analyses in JASP[1], a user-friendly and freely available software package. JASP was intentionally designed to make Bayesian methods more accessible to a wider audience of researchers, and several scientific fields are already starting to embrace the use of the JASP platform, such as numerical cognition (Faulkenberry et al., 2020), psychiatry (Quintana & Williams, 2018), and biomedicine (Kelter, 2020).

By virtue of JASP software being free to download and actively updated, with a large-and-expanding user community, JASP makes it much easier for academics and practitioners alike to conduct basic Bayesian analyses (JASP Team, 2023; Love et al., 2019; Marsman & Wagenmakers, 2017) such as for linear regression (van den Bergh et al., 2021) and ANOVA (van den Bergh et al., 2019). Moreover, JASP can also conduct traditional frequentist (NHST) statistical analyses, which allows for an easy side-to-side comparison with Bayesian results, as we will demonstrate here. This way, organizational researchers can learn basic Bayesian analysis with the comfort that comes with greater understanding.

It is also important to note how JASP software also promotes various aspects of transparency and reproducibility, in support of open science (JASP Team, 2023). First, JASP is free to download and use on any operating system (at https://jasp-stats.org/), which helps remove financial and technical barriers to cultivating a more inclusive and diverse research community. Second, JASP also promotes data accessibility by including direct downloading and uploading to the Open Science Framework (OSF; https://osf.io/; Wagenmakers et al., 2018), which is increasingly used by organizational researchers. Third, data and analyses can both be saved within one integrated JASP file, which documents and consolidates one's analysis workflow. And fourth, tables and figures can be copied and pasted directly into the desired document in near-APA format, which avoids transcription errors.

Many of the capabilities of JASP sound similar to those of R (R Core Team, 2023). R is also an open-source and platform-friendly computer software program, where users can choose among literally thousands of R packages dedicated to data wrangling, statistical analysis, and data visualization. Often, packages are installed from the Comprehensive R Archive Network (CRAN), which vets packages before they are posted to a set of volunteer mirrored distribution sites, but other packages are also available from GitHub and personal websites. Although R has a steeper learning curve than JASP by requiring users to learn a coding language to perform analyses and packages, R provides greater flexibility and control in building Bayesian models as a result. Conversely, JASP is easier for learning basic Bayesian analyses via drop-down menus and an accessible user interface, which provides more users easier access to obtaining a basic understanding of Bayesian analyses and their frequentist counterparts: e.g., *t*-tests, ANOVAs, mixed models, regressions, and frequencies (JASP Team, 2023). Importantly, the basic understanding from JASP provides a smoother transition into using R for conducting more thorough and complex Bayesian analyses. Also, as of January 2023, JASP (version 0.17) now includes syntax mode which allows users to export R code of their JASP analysis, much like jamovi, another open-source statistical software program (JASP, 2023; The jamovi project, 2023).

Supplementing these default menus, the Bayesian capabilities of JASP can be extended: e.g., by importing a Learn Bayes module, and by entering popular JAGS (Just Another Gibbs Sampler) code to run, evaluate, and plot customized analyses. These Bayesian options available to users of JASP continue to be expanded and refined over time by the JASP software team, as informed by the aforementioned JASP user community (e.g., https://github.com/jasp-stats/jasp-issues/issues). With this being said, JASP is still actively developed, which means that it is still lacking some features (e.g., effect size for Bayesian ANOVA, prior specification for Bayesian linear mixed models).

Although fitting much more complex models to data is a key benefit of adopting Bayesian analysis (Best et al., 1996), our primary goal for this paper is to

[1] JASP stands for Jeffreys's Amazing Statistics Program, a humorous homage to Sir Harold Jeffreys, author of the *Theory of Probability* (1939) and one historically credited for helping to revive Bayesian inference.

present the mechanics and benefits of basic Bayesian modeling to an organizational research audience, thereby providing a solid conceptual and practical foundation from which the reader can then seek out and understand more complex Bayesian modeling more confidently.

The Bayesian Basics Using JASP

In this section, we will use JASP to conduct a multiple linear regression and a linear mixed model. For each method, we will first apply the traditional frequentist (NHST) approach with which readers are familiar, using that for a useful comparison when next demonstrating a Bayesian analysis as its counterpart. For each of these Bayesian analyses in JASP, we will provide organizational examples, so that reporting and interpreting results are domain-relevant, and we will supply straightforward procedures for translating key elements of JASP output into the writeup of Bayesian organizational research results for publication.

Bayes Theorem

Before describing the tutorial examples, it is important to first discuss what underlies a Bayesian analysis: Bayes Theorem. Bayes Theorem has three essential components that support the process of updating beliefs: the prior, the likelihood, and the posterior (see Table 1 for some definitions of some common Bayesian terms). Given that H = a hypothesized model and E = the evidence (observed data), Bayes Theorem mathematically combines these three aforementioned elements:

$$P(H|E) = \frac{P(H) * P(E|H)}{P(E)}$$

The *prior distribution*, or P(H) in the numerator, is a probability distribution that operationalizes what we know about possible values of the parameter of interest before ever observing and modeling data. Once the data are observed, the *likelihood*, expressed mathematically as P(E | H) in the numerator, describes how likely it is that we observed the data, given our hypothesized model. As you can see, the likelihood is multiplied by the prior in the numerator of Bayes Theorem, meaning that the likelihood of observing our data is weighted by our prior assumptions going into the analysis.

This numerator is simply divided by P(E) in the denominator, which simply reflects all possibilities of the hypothesized model [where P(E) = P(H) * P(E | H) + P(¬H) * P(E | ¬H), and ¬ means 'not']. This division normalizes the result to yield the *posterior distribution*, a probability distribution that tells us how likely our hypothesized model is based on the data, or P(H | E). Note that the posterior distribution will always be situated between the prior distribution and the likelihood, such that the more data one has, the closer the posterior will be located to the likelihood than to the prior. On the other hand, the less data one has, the posterior distribution will be shifted closer to the prior, because the data (likelihood) will have less influence. Bayes Theorem ensures that this calibration of the posterior distribution toward the data versus toward prior is mathematically principled.

In addition to covering the fundamentals of Bayes Theorem here, we also need to introduce the concept of the *Bayes factor* before turning to our organizational examples. The Bayes factor (BF) provides the relative weight of the data (evidence) supporting one model versus another, such as the null versus alternative hypothesis, using Bayes Theorem. Using JASP, the "Bayes Factor" option asks the user to select between "BF_{10}," "BF_{01}," and "Log(BF_{10})." Regarding these subscripts, note that (a) 0 and 1 refer to the null and alternative hypotheses respectively, and (b) the number listed first is in the numerator, with the second number in the denominator. For example, a BF_{10} of 5 would mean that the data are 5 times more likely under the alternative hypothesis than the null hypothesis. Or conversely, a BF_{01} of 5 would mean that the data are 5 times more likely under the null hypothesis than the alternative. Note that BF_{10} and BF_{01} are inverses. For example, the BF_{10} of 5 is equivalent to a BF_{01} of ⅕; these values would have the same interpretation. Note that if $BF_{10} = 1$, then also $BF_{01} = 1$, meaning that both hypotheses are equally likely.

The Bayes factor departs from the frequentist (NHST) approach because we are comparing two models; we are not simply rejecting or failing to reject the null hypothesis. Generally speaking, an organizational researcher might be more likely to report a BF_{10} value, given that the alternative hypothesis is usually the research hypothesis of interest. The Log(BF_{10}) metric can also be useful when plotting and comparing BF values, because the natural log of the Bayes factor makes values symmetric

Table 1. Definitions of Common Bayesian Terminology

	Definition
Prior	A probability distribution that operationalizes what we know about possible values of the parameter of interest before ever observing and modeling data
Likelihood	A probability distribution that describes how likely it is that we observed the data given our hypothesized model
Posterior	A probability distribution that updates our prior distribution, based on the likelihood; it tells us what we know about possible values of the parameter of interest after observing and modeling the data
Bayes factor	The relative weight of the data (evidence) supporting one model versus another
Credible interval	An interval that contains a specific percentage of the most probable parameter values (usually this is 95%)

around zero: e.g., we noted that a BF_{10} of 5 is equivalent to a BF_{01} of ⅕; however, if you take the natural log of both, they become equidistant from zero, at 1.61 and -1.61, respectively.

In the frequentist (NHST) paradigm, we typically consider *p* values being statistically significant if they are less than a predetermined alpha level (e.g., .05 or .01), and they are not statistically significant otherwise. Within the Bayesian paradigm, the terminology for the size of the BF is attached to *multiple* cutoffs, to operationalize how the weight of evidence for one hypothesis versus another lies along a continuum (versus dichotomous language significance/non-significance attached to *p* values). One reasonable set of cutoffs provided by Jeffreys (1961), and later revised by Lee and Wagenmakers (2013), proposed the following system of labels for determining the strength of BF_{10}: a range of 0-3 provides anecdotal evidence, 3-10 provides moderate evidence, 10-30 provides strong evidence, 30-100 provides very strong evidence, and > 100 provides extreme evidence (see Table 2). However, just like the phrase *statistical significance* used in the frequentist (NHST) paradigm, we acknowledge that this language for Bayesian testing may still contribute to the habit of applying labels mechanically without a closer consideration of practical significance (such as examining the size of *d*-values or correlations in frequentist statistics). Practical significance is an issue that informs statistical findings, but also lies outside of them by considering the nature of the research question, the appropriateness of the models being compared, the meaning of the descriptive statistics from which inferences are made, and so on.

Multiple Linear Regression Analysis

For the multiple linear regression analyses, we used modified and anonymized data we received from an organization. This dataset has a total of 271 participants that were assessed on various skills and performance standards. Our research question is: Do employee computer skills and the ability to meet performance goals predict overall performance? Our null hypothesis is that these variables, when modeled together in a linear regression, do not predict overall performance, and our alternative hypothesis is that together they predict overall performance. We checked that the assumptions of a multiple linear regression analysis were generally met (e.g., based on standards for assessing linearity, independence of observations, homoscedasticity, and normal distribution of the

Table 2. Evidence Provided by the Bayes Factor

BF_{01} ranges	Evidence[a]	BF_{10} ranges
$\frac{1}{3}$ - 1	Anecdotal	1-3
$\frac{1}{3}$ - $\frac{1}{10}$	Moderate	3-10
$\frac{1}{10}$ - $\frac{1}{30}$	Strong	10-30
$\frac{1}{30}$ - $\frac{1}{100}$	Very Strong	30-100
$< \frac{1}{100}$	Extreme	> 100+

Note. BF_{01} is the reciprocal of BF_{10}, such that BF_{01} (leftmost column) indicates the amount of evidence supporting the null over the alternative hypothesis explaining the data, and conversely, BF_{10} indicates the amount of evidence supporting the alternative hypothesis over the null hypothesis explaining the data. As can be seen, values that depart from BF = 1 indicate stronger support for one model over another.

[a] Adapted from Lee and Wagenmakers's (2013) revision of Jeffreys's (1961) classification.

Table 3. Descriptive Statistics and Correlations: Computer Skills, Meet Performance Goals, & Overall Performance

Variables	*N*	Mean (*SD*)	1.	2.	3.
1. Computer Skills	270	3.77 (0.91)	-		
2. Meets Performance Goals	271	3.15 (0.79)	.51[a]	-	
3. Overall Performance	271	3.73 (0.85)	.58[a]	.75[a]	-

Note. [a] $BF_{10} > 100$

residuals). We also chose not to remove any statistical outliers in this dataset (10 were identified in univariate outlier analysis by being outside 1.5 interquartile ranges from the 25th and 75th quartiles), because they represent natural variations in the data (e.g., employees who were rated a 1 out of 5 on Meets Performance Goals). But as a side note, the parameter estimates remained relatively unchanged when these outliers were removed. Table 3 provides descriptive statistics for the variables included in the model. All JASP files used in analysis can be found on OSF (https://osf.io/3jhx6/).

Frequentist (NHST) Analysis

Set-up. To compute a classical regression, we select "Regression" – "Classical" – "Linear Regression." Our dependent variable is overall performance, and our covariates are computer skills and ability to meet performance goals. We dragged and dropped "Overall Performance" to the "Dependent Variable" box and moved "Computer Skills" and "Meets Performance Goals" to the "Covariates" box. JASP automatically computes the model and outputs a "Model Summary" table, an "ANOVA" table, and a "Coefficients" table. To check for multicollinearity, a user can select

"Statistics" – "Collinearity diagnostics" to add "Tolerance" and "VIF" measures to the "Coefficients" table.

Analysis. Figure 1 shows the JASP output. The "Model Summary" as shown in Table 4 contains two models: the null and the alternative. For our purposes, we will focus on the alternative model. The table contains three columns: R^2, R^2 Adjusted, and RMSE.

We observed an adjusted R^2 of .61, meaning that 61% of the variation in overall performance is accounted for by computer skills and meeting performance goals taken together. R^2 adjusted is the value that is typically reported; it applies a statistical penalty to the R^2 value for the number of predictors in the model, in an attempt to adjust for overfitting that can happen when too many predictors are entered into the model, relative to the sample size.

RMSE (root mean squared error) tells us how far the predicted values are from the observed values. For example, we observe H_1 having a RMSE of 0.53, which is the average deviation of the model's predicted value of overall performance from the observed value. Lower RMSE values indicate a better model fit. Table 5 contains the test of the model against the null hypothesis. The first column contains the alternative "Model," which is summarized in column two as the regression, residuals, and total. The third column contains the "Sum of the Squares," which is a function of the deviation of observed values from the mean. The third column contains the model degrees of freedom. The fourth column contains the "Mean Square," which is the sum of the squares divided by its corresponding degrees of freedom. The fifth and sixth columns contain the "*F*" statistic and "*p*" value, respectively. Of course, the *p* value is based on the obtained *F* statistic to determine whether the null hypothesis is statistically significant and can therefore be rejected in support of the alternative hypothesis.

The "Coefficients" table, as shown in Table 6, contains the "Model" specification, "Unstandardized" coefficient, "Standard Error," "Standardized" coefficient, "*t*," and "*p*" of the coefficients, "Computer Skills" and "Meets Performance Goals." This table is common to any frequentist regression analysis.

Figure 1. Screenshot of JASP Output for a Frequentist Multiple Linear Regression

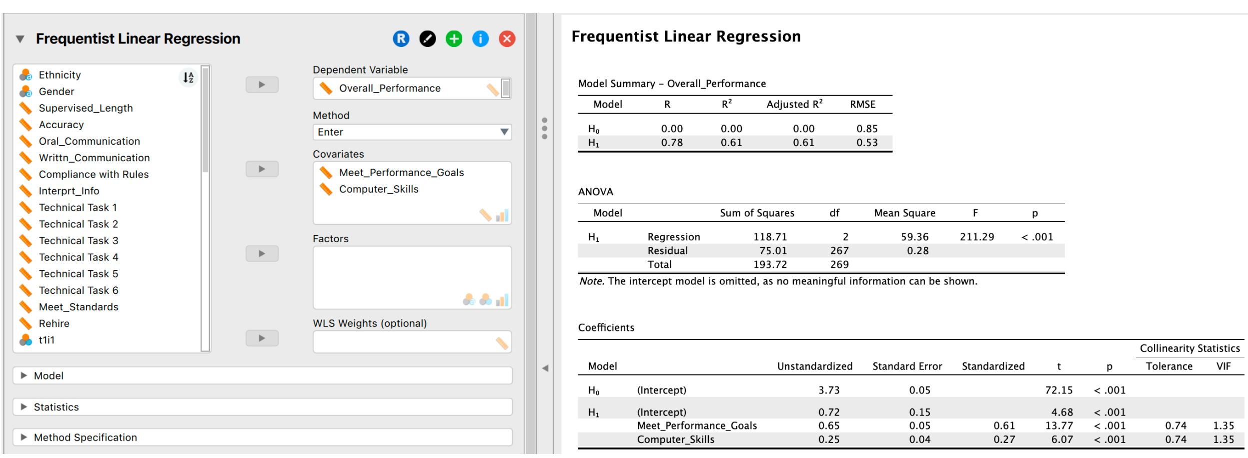


Table 4. Model Summary for Multiple Linear Regression: Overall Performance

Model	R^2	Adjusted R^2	RMSE
H_0	.00	.00	0.85
H_1	.61	.61	0.53

Note. Total N = 271.

Table 5. Test of the Frequentist Multiple Linear Regression Model: Overall Performance

Model		Sum of Squares	*df*	Mean Square	*F*	*p*
H_1	Regression	118.71	2	59.36	211.29	< .001
	Residual	75.01	267	0.28		
	Total	193.72	269			

Note. Total N = 271. *df* = Degrees of freedom.

Interpretation. A template for reporting these frequentist findings might be: "We conducted a linear regression to examine if overall performance is predicted by computer skills and meeting performance goals. We observed that computer skills and meeting performance goals together explained a statistically and practically significant portion of the variance of overall performance, R^2 = .61, $F(2, 267) = 211.29$, $p < .001$. The slope coefficient for computer skills, $\beta_1 = 0.25$, $t(267) = 6.07$, $p < .001$, and the slope coefficient for meeting performance goals, $\beta_2 = 0.65$, $t(267) = 13.77$, $p < .001$, are both positive and statistically significant, indicating that they each uniquely and positively predict overall job performance." Note that the intercept is arbitrary and should be tabled, but it is only interpreted if the value of 0 for all predictors is meaningful (and here it is not).

Bayesian Analysis

Set-up. We selected "Regression" – "Bayesian" – "Linear Regression." As before, we dragged and dropped overall performance, computer skills, and meeting performance goals to the appropriate boxes. We used the default prior on the models, a beta-binomial prior with ($\alpha = \beta = 1$), denoted as (a = b = 1) in JASP, which specifies priors on models by the number of predictors included (for more information on priors in linear regressions, see Consonni et al., 2018; Liang et al., 2008). To compare our models to the null model, be sure to select "Compare to null" under "Order." Finally, we selected "Output" – "Posterior Summary" – and "Plot of coefficients."

Analysis. JASP automatically outputs a table with the model in the right panel as shown in Figure 2. Table 7 contains the model comparison and Table 8 contains the posterior summary statistics. We can use the first table to determine which model best explains the data and the second table to understand the relative ability of each model to explain the data. JASP automatically orders the models from best to worst in terms of predicting the observed outcome. The default setting for regression analysis in JASP uses the same algorithm as the Bayesian Adaptive Sampling (BAS) package in R which applies an adaptive sampling algorithm that is more appropriate and efficient for sampling a larger model space (see Clyde et al., 2011; Clyde, 2018).

Starting the first column of Table 7, P(M) is the prior on the models. The P(M) value for the two-variable model equals .33. The P(M) for the one-variable models also equals .33, but then gets split equally across the two models, receiving P(M) values of .17 each. This type of weighting based on the number of predictors in the model (rather than just weighting all individual models equally) can be done with a beta-binomial prior (e.g., van den Bergh et al., 2021) which corrects for multiplicity (Scott & Berger, 2010). The prior settings in JASP allow users to modify the prior probabilities of the models or change the prior applied to the parameters (i.e., the regression coefficients) themselves. In the present case, the default prior on the regression coefficients is a Jeffreys-Zellner-Siow (JZS) prior, which applies a Cauchy distribution to the effect size (μ/σ) and Jeffreys's prior (Jeffreys, 1961) to the variance (Rouder et al., 2009; Rouder & Morey, 2012; Wetzels et al., 2009). A Cauchy distribution is equivalent to a Student's t distribution with one degree of freedom and looks similar to a normal distribution, but has thicker tails. Rouder and Morey (2012) call the JZS prior a *default* prior because of the three favorable properties of the resulting Bayes factor: location-scale invariance (e.g., changing units of measurement should not change the analysis), consistency (i.e., the BF should favor models that are actually better explained by the data), and information-consistency (i.e., as R^2 approaches one, the BF in support of the alternative hypothesis approaches infinity).

Other prior options in JASP include: AIC, BIC, EB-Global, EB-Local, g-prior, hyper-g, hyper-glaplace, and hyper-g-n (see Liang et al., 2008). This paper does not cover prior selection in detail because we are assuming that we do not have any prior knowledge about the parameter space, and therefore

Table 6. Multiple Linear Regression Model Coefficients: Overall Performance

Model		Unstandardized	Standard Error	Standardized	*t*	*p*
H_0	(Intercept)	3.73	0.05		72.15	< .001
H_1	(Intercept)	0.72	0.15		4.68	< .001
	Meet Performance Goals	0.65	0.05	0.61	13.77	< .001
	Computer Skills	0.25	0.04	0.27	6.07	< .001

Note. Total *N* = 271. Unstandardized = Unstandardized mean coefficient. Standardized = Standardized mean coefficient.

Figure 2. Screenshot of JASP Output for a Bayesian Multiple Linear Regression

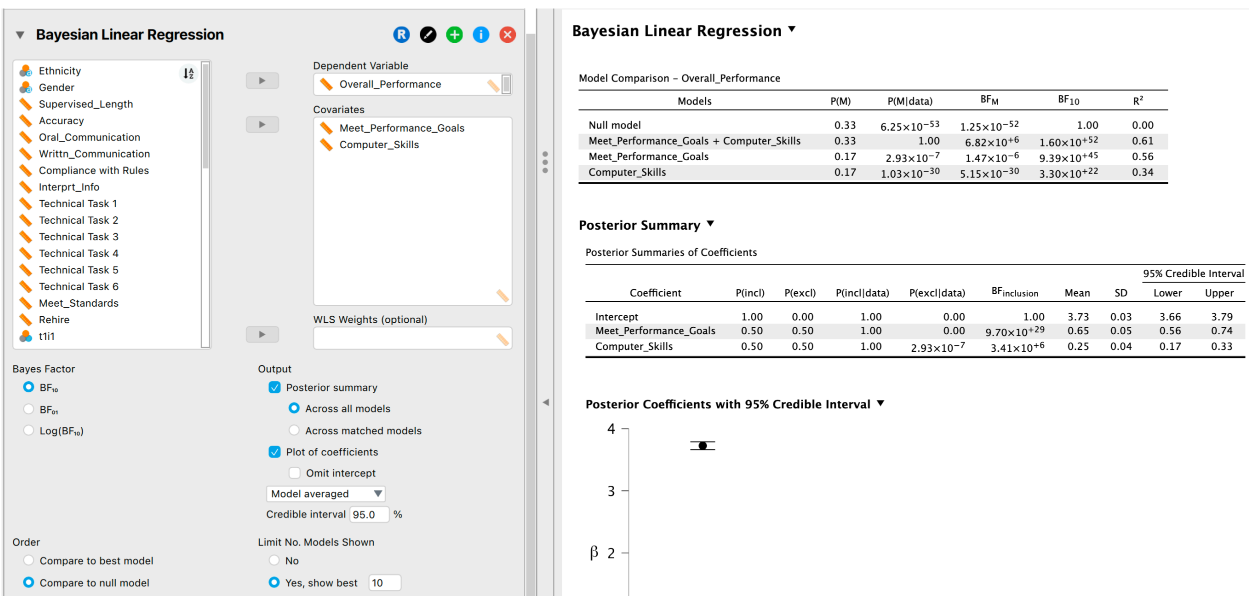


Table 7. Bayesian Model Comparison: Overall Performance

Models	P(M)	P(M|data)	BF_M	BF_{10}	R^2
Null model	.33	6.25e -53	1.25e -52	1.00	0
Computer Skills + Meet Performance Goals	.33	1.00	6.82e +6	1.60e +52	.61
Meet Performance Goals	.17	2.93e -7	1.47e -6	9.39e +45	.56
Computer Skills	.17	1.03e -30	5.15e -30	3.30e +22	.34

Note. Total *N* = 271. P(M) = Prior. P(M|Data) = Posterior probabilities. BF_M = Bayes factor of model odds. BF_{10} = Bayes factor of model relative to null. R^2 = Proportion of explained variance in the outcome.

Table 8. Bayesian Posterior Summaries of Coefficients: Overall Performance

Coefficient	P(incl)	P(excl)	P(incl|data)	P(excl|data)	BF $_{inclusion}$	Mean Coeff.	*SD*
Intercept	1.00	0	1.00	0	1.00	3.73	0.03
Meet Performance Goals	.50	.50	1.00	0	9.70e+29	0.65	0.05
Computer Skills	.50	.50	1.00	2.93e -7	3.41e+6	0.25	0.04

Note. Total N = 271. P(incl) = Prior probability of inclusion in the model. P(excl) = Prior probability of leaving out of the model. P(incl|data) = Posterior probability. P(excl|data) = Posterior probability. $BF_{inclusion}$ = Bayes factor relative to null. Mean Coeff = Mean coefficient estimate. SD = Standard deviation.

using default priors are adequate for our purposes. More generally, the guardrails provided by JASP guide users in making reasonable and practical decisions within the software. For users who want to use informative priors (e.g., as informed by other empirical findings or by expert judgements), we encourage them to use R and the rich array of available Bayesian resources (e.g., Gelman et al., 2015; Gronau et al., 2020; Kruschke, 2014; McElreath, 2020; Stefan et al., 2019; van de Schoot et al., 2018). We did however examine sensitivity to the prior by running the same model while varying the scale of the JZS prior (select "Advanced Options" - "JZS" - "r scale") to a wider scale (i.e., increasing *r*); the best performing model and pattern of $BF_{inclusion}$ probabilities remained the same. Checking prior sensitivity is essential for examining the impact of the prior on posterior inference (e.g., Depaoli, 2020; van Erp et al., 2017).

The notation P(M|data) indicates posterior probabilities, in other words, the probability of the models after observing data. These are ordered from greatest to least, with the top model "Computer Skills + Meets Performance Goals" having a posterior probability that is nearly 1. The column of P(M|data) values should add to 1 because they are probabilities, so this model is much better at explaining the data than the others. Between these first two columns, we can see how the priors change once data is observed to output a posterior probability. To recap the process presented in this table, we can think of this as (1) believing that the size of the models is equally likely, (2) applying default priors to the regression coefficients, (3) viewing the data, (4) updating our beliefs to understand that the "Computer Skills + Meets Performance Goals" model fits the data the best.

In the third column, BF(M) is the Bayes factor of the model odds after observing data. The odds of the "Computer Skills + Meets Performance Goals" before observing the data were .33 / (.33 + .17 + .17) = .50. Looking at the P(M|data) column, we can see that after observing the data we updated our prior from .5 to nearly 1 (the probability is so large that it rounds to 1), indicating that the probability increased after observing data. To get the BF_M, we divide the posterior odds by the prior odds. The BF_{10} column has the Bayes factors of each model relative to the null model. For example, the "Computer Skills + Meets Performance Goals" model provides the most evidence and is 1.60e+52 times better at explaining the data, relative to the null model (this number is enormous, compared with BF typically found and reflected in Table 2).

The posterior summary in Table 8, contains the model's coefficients. Starting with the seventh column, we can find our estimates for the values of the intercept and slopes and the eighth column contains the standard deviation of the mean estimates. Going back to the first column, we can find the name of the coefficient. The second and third columns are "P(inclu)," which is the prior probability of including the coefficient in the model, and "P(exclu)" is the prior probability of excluding the coefficient, since we used a beta-binomial prior ($\alpha = \beta = 1$), on the models, each coefficient is initially presumed to be equally likely to be included. Again, the prior on the regression coefficients themselves is a default, JZS prior. The fourth column, "P(inclu|data)," is the posterior probability, that is the probability of including the coefficient after observing the data and the fifth column "P(exclu|data)," is the posterior probability of excluding the coefficient after observing the data. Finally, the sixth column "$BF_{inclusion}$" tells us the Bayes factor of including the coefficient in the model. Figure 2 also shows the 95% credible intervals (CrIs) of the mean posterior coefficients which tells us the location of the most probable posterior coefficients.

Interpretation. For guidance regarding what to report, see Depaoli and van de Schoot (2017), Kruschke (2021), van Doorn et al., (2021), and van de Schoot et al. (2021). To report our findings, we could state the following: "We conducted a Bayesian linear regression to examine if computer skills and meeting performance goals predicted overall performance. We used a beta-binomial prior ($\alpha = \beta = 1$) on the models, which assumes all model sizes are equally likely to explain the data, and we applied JASP's default prior, a JZS prior (r = .354), on the regression coefficients. JASP sampled the model space using the Bayesian Adaptive Sampling (BAS) package from R (Clyde, 2022), which can be more appropriate than MCMC methods when engaging in larger-scale variable selection. We observed that the model composed of both computer skills and meeting performance goals explained a large amount of the variance in overall performance, R^2 = .61 (BF_{10} > 100). The data are in extreme favor of the alternative model (BF_{10} > 100) compared to the null model. Furthermore, the data extremely support including both computer skills ($BF_{inclusion}$ > 100) and meeting performance goals ($BF_{inclusion}$ > 100) when predicting overall performance." If we are interested in what the probable parameter values are for each parameter in the model, we can also report the mean and the corresponding 95% CrI for each coefficient.

Linear Mixed Models (Multilevel Modeling)

This second more advanced example follows the frequentist multilevel modeling example provided by Bliese (2022, data from Bliese & Halverson, 1996). This multilevel model is also called a *linear mixed regression*, where here we are examining whether soldier well-being across different work groups (i.e., 99 Army companies with an average of about 75 soldiers) are predicted by both individual-level work demands (i.e., average individual work hours in a day, our level-one variable) and group-level work demands (i.e., average work hours at the group level; our level-two variable). This model is conceptually useful because we can determine the extent to which soldier well-being is influenced by (a) work at the overall group level (i.e., generally, the more the group works, the lower the individual well-being), (b) independent of the group, the time that each soldier works (i.e., generally, soldiers who work more than the group average have lower well-being). We first estimate the parameters of a multilevel model that allows the intercept for the relationship between work demands and well-being to vary across groups (i.e., estimating random intercepts variance), while estimating the slope as a single constant value across groups. Next, we extend this multilevel model by allowing both intercepts and slopes to vary across groups (i.e., estimating random intercepts variance, random slopes variance, and random intercept-slope covariance). In checking model assumptions, we observed a linear relationship between the predictors and well-being, observed homoscedasticity of variances, and the residuals appeared to be normally distributed.

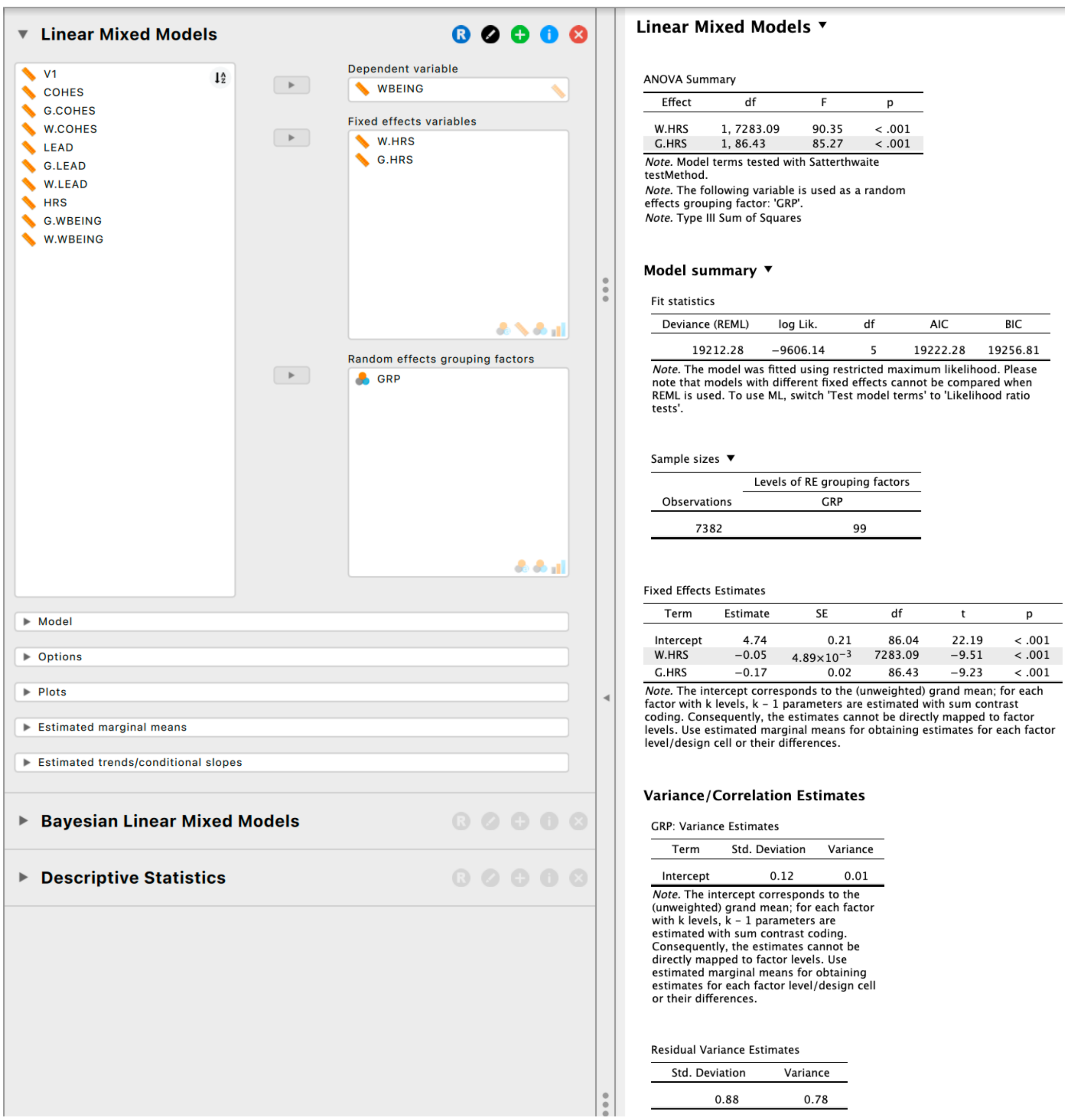
The dataset contains the two main variables that address the research question above: Well-Being (WBEING) as our dependent variable, with Group Hours (G.HRS) reflecting the average daily work hours for each group, and Work Hours (W.HRS) reflecting the average daily work hours for each soldier. Note that W.HRS is expressed as a deviation of soldiers' work hours above or below their respective group mean (e.g., if a soldier works 7 hours, and G.HRS = 5, then W.HRS = 2). This makes between-group and within-group effects of work hours completely independent, and thus they are interpreted independently below. In our program code, note that set the random number seed to 1 in both examples for reproducibility purposes.

Frequentist – Varying Intercepts

Setup. To compute a frequentist linear mixed model in JASP, we start by selecting "Mixed Models" – "Classical" – "Linear Mixed Models" from the dropdown menu. We then dragged and dropped "WBEING" to the dependent variables box, "G.HRS" and "W.HRS" to the fixed variables box, and "GRP" to the random effects grouping variables box. Next, under "Random effects," we only checked the box next to "Intercept" because we are interested in how the intercept varies by each group, keeping the slopes for W.HRS constant across groups (and G.HRS has a single slope across all groups), see Figure 3. Note that JASP defaults to a varying intercepts and slopes model, but in this first model, we only want to vary the intercepts. We can also select "Options" – "Differences from intercept," "Model summary," "Fixed effects estimates," and "Variance/correlation estimates" to examine model fit further.

Analysis. JASP automatically populates a table in the right pane with ANOVA summary statistics and under the "Options" tab, we can also select "Model summary," "Fixed effects estimates," and "Variance/correlation Estimates" which populates additional tables with fit statistics for each predictor, fixed effects estimates (intercept and slope coefficients; see Table 9), group variance estimates, group correlation estimates, and residual variance estimates. The intercepts for W.HRS have an estimated standard deviation of 0.12 and the residual (unexplained) standard deviation of the overall model is 0.88.

Figure 3. Screenshot of JASP Output for a Frequentist Linear Mixed Model



Interpretation. We can interpret the results as: "We observed a negative and statistically significant between-group effect of work hours on well-being, $\beta_2 = -0.17$, $t(86.43) = -9.23$, $p < .001$, indicating that the total group effect is reliably different from zero (see Table 9). Moreover, for every hour increase at the group level, well-being is predicted to decrease by -0.17 points. Additionally, we observed a negative and statistically significant within-group effect of work hours on well-being, $\beta_1 = -0.05$, $t(7293.09) = -9.51$, $p < .001$. Thus, within each group, for every one-hour increase in individual working hours, predicted well-being ratings decrease by .05 points. Finally, the intercepts for within-group work hours have an estimated standard deviation of 0.12 (see Table 10)." Note that we group-mean centered work hours and reintroduced the means as a between-group variable. Therefore, the between-group coefficient and within-group coefficient for work hours are uncorrelated, and there are no other predictors in the model. In this special case, the between- and within-group variables capture distinct sources of variation in the outcome and can be interpreted independently.

Frequentist – Varying Intercepts and Varying Slopes

Setup. To compute a frequentist linear mixed effects model with varying intercepts and slopes, we follow the same setup as for the varying intercepts models, but instead under "Random effects," we now check the boxes next to "Intercept," "G.HRS," and "W.HRS" to examine a model where the between-group slope for group hours is modeled, and the slopes for individual hours of work well-being are allowed to vary within each group.

Analysis. JASP automatically populates similar tables as we observed in the varying intercepts model. The coefficients are highly comparable to the

Table 9. Predicting Soldier Well-Being:
Frequentist Fixed Effects Estimates for the Varying Intercepts Model

Term	Estimate	SE	*df*	*t*	*p*
Intercept	4.74	.21	86.04	22.19	< .001
W.HRS	-0.05	.01	7293.09	-9.51	< .001
G.HRS	-0.17	.02	86.43	-9.23	< .001

Note. Total $N = 7{,}382$. W.HRS = group-mean centered hours (level one variable). G.HRS = group hours (level two variable).

Table 10. Predicting Soldier Well-Being:
Frequentist Random Effects Estimates for the Varying Intercepts Model

Level	Estimate
Level 2 (between)	
Intercept	0.12
Level 1 (within)	
Residual	0.88

Note. Total $N = 7{,}382$. Estimates are standard deviations.

varying intercepts model (see Table 11). The estimated standard deviation of the intercept of group work demands is 0.12 and the standard deviation of the slopes is 0.02. The correlation between the intercepts and slopes is -.29 (see Table 12). The unexplained standard deviation of within group work hours is 0.88.

It is useful to explain briefly the fit statistics provided by JASP (i.e., Deviance REML, Log Likelihood, AIC, and BIC) for comparing each model. These criteria generally weigh model prediction versus model complexity to select the "best" model. The Deviance REML index accounts for fixed effects to give an unbiased estimate of the random-effects variances, but we cannot use the index to compare across models with differing fixed effects. The Log Likelihood index does not penalize more complex models, where values closer to zero indicate better model fit. The Akaike Information Criterion (AIC; Akaike, 1974) and Bayesian Information Criterion (BIC; Schwarz, 1978) penalize more complex models, and smaller values indicate better model fit. The primary difference between AIC and BIC lies in their underlying assumptions about models; AIC estimates the Kullback–Leibler (K-L) divergence between the unknown, true model and the model being examined, whereas BIC is based on BFs and assumes that as N increases the true model is chosen with increasing probability (see Vrieze, 2012). We choose to use AIC, given one cannot justify a 'true' model among the candidates being compared. The AIC values for the

Table 11. Predicting Soldier Well-Being:
Frequentist Fixed Effects Estimates for the Varying Intercepts + Slopes Model

Term	Estimate	*SE*	*df*	*t*	*p*
Intercept	4.73	0.21	86.19	22.21	< .001
W.HRS	-0.05	0.01	87.69	-8.36	< .001
G.HRS	-0.17	0.02	86.59	-9.20	< .001

Note. Total N = 7,382. W.HRS = group-mean centered hours (level one variable). G.HRS = group hours (level two variable).

Table 12. Predicting Soldier Well-Being:
Frequentist Random Effects Estimates for the Varying Intercepts + Slopes Model

Level	Estimate	Intercept	W.HRS
Level 2 (between)			
Intercept	0.12	–	
W.HRS	0.02	-.29	–
Level 1 (within)			
Residual	0.88		

Note. Total N = 7,382. Estimates are standard deviations, and a correlation in the square matrix.

varying intercepts is 19222.3 and the varying intercept and slopes model is 19220.9.

Interpretation. The results in Table 11 and Table 12 can be interpreted similarly to the varying intercepts model. The main difference is that we can now assess if a model that lets both the intercepts and slopes vary for each group improves model fit over the model with varying intercepts but a fixed slope. We can interpret the results as: "We observed a negative and statistically significant between-group effect of work hours on well-being, β_2 = -0.17, $t(86.59)$ = -9.20, $p < .001$, indicating that the total group effect is reliably different from zero (see Table 11). Moreover, for every hour increase at the group level, well-being is predicted to decrease by -.17 points. Additionally, we observed a negative and statistically significant within-group effect of work hours on well-being, β_1 = -0.05, $t(87.69)$ = -8.36, $p < .001$. Thus, for every one-hour increase in individual average working hours, predicted well-being ratings decrease by .05 points. The random effect of within-person work hours is small (SD = 0.02), which suggests that there is little variation between soldiers (see Table 12). The AIC is 2 lower for the varying intercepts and slopes model, which means that this model is favored. Changes greater or equal to 2 (but less than 4) are considered to reflect substantial support of a model (Burnham & Anderson, 2004)."

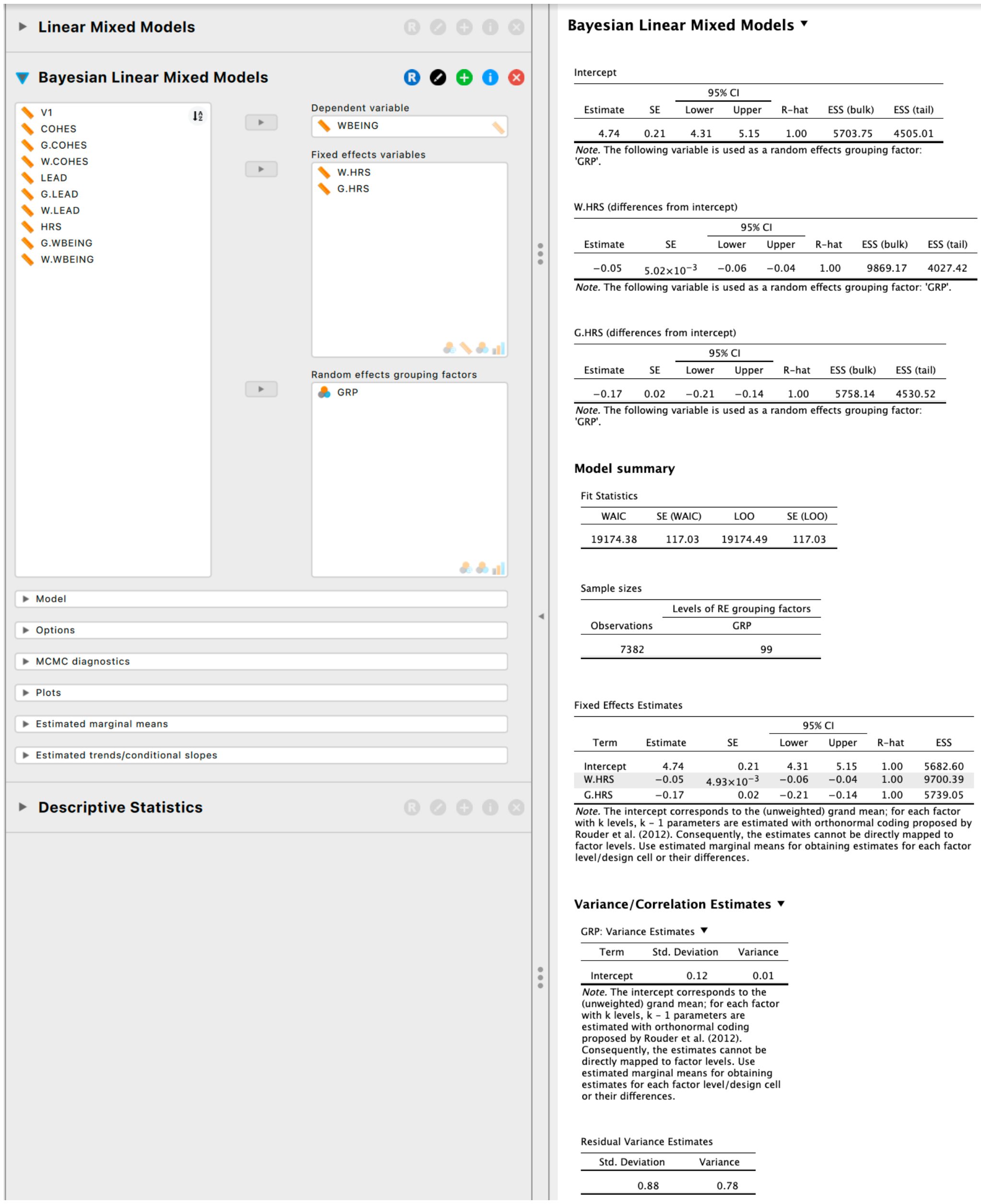
Bayesian – Varying Intercepts

Set up. To compute a Bayesian linear mixed model in JASP, we started by selecting "Mixed Models" – "Bayesian" – "Linear Mixed Models" from the dropdown menu. We then dragged and dropped the same variables as in the frequentist example (i.e., "WBEING" to the dependent variables box, "W.HRS" and "G.HRS" to the fixed variables box, and "GRP" to the random effects grouping variables box, see Figure 4). We used JASP's default settings, which is 2,000 posterior samples as burn-in (the number of warmup samples that are discarded), 4,000 actual posterior samples for each chain (samples used in analysis), 3 chains (the number of chains that completed the posterior sampling), .8 adapt delta (the average target acceptance rate for whether a proposed value goes into the chain), and 10 maximum tree depth (the greatest number of evaluations per sample; if this is ever exceeded, then the sampling steps are too small). These default settings can be modified if the chains do not 'mix' or converge adequately (this can be assessed by examining plots of the MCMC chains, see Appendix A and via the R-hat statistic, see Figure 4). For example, if there are *divergent transitions* after the warmup (i.e., the Hamiltonian Monte Carlo sampler goes off course from the expected path, see Stan Development Team, 2022a), a user is recommended to try increasing the adapt delta value to .90 or .95, and if there are warnings about R-hat and effective sample size (ESS) values (the number of samples in different parts of the posterior distribution) the user is recommended to try increasing the number of burn-in samples or sampling iterations (see Stan Development Team, 2022b for more ways to diagnose convergence issues).

Finally, under "Options" – "Show" select "Differences from intercept," "Model summary," "Fixed effects estimates," and "Variance/correlation estimates."

Analysis. JASP automatically populates a table (Table 13) with the intercept and each of the slope coefficients. Standard error "SE", "Lower" and "Upper" 95% CrI, "R-hat," "ESS (bulk)," and "ESS (tail)." Focusing on these latter three new terms, an R-hat (i.e., scale reduction factor; Gelman & Rubin, 1992) tells us if the algorithm has converged, values at 1 indicate convergence of the MCMC chains, whereas values greater than 1.1 suggest that greater posterior distribution sampling is needed (e.g., increasing the number of iterations). ESS, effective sample size, (bulk) is the number of observations at the center of the distribution, where smaller numbers represent imprecision in the parameter, and the ESS (tail) specifies the number of observations in the tails of the distribution, where smaller numbers represent imprecision in the CrI bounds (JASP, 2023). Generally, 1,000 samples in the bulk and tail of the distribution are sufficient (Zitzmann & Hecht, 2019). See Appendix A for the mixing of the two MCMC chains for both variables, and see Appendix B for autocorrelation plots, which suggest that MCMC estimates are sufficiently uncorrelated. There are currently no options for modifying the prior; JASP automatically applies the default, auto-scaled prior used in the `rstanarm` R package for linear mixed models, which applies an exponential distribution on the standard deviation of the random effect and an independent normal distribution centered at 0 and

Figure 4. Screenshot of JASP Output for a Bayesian Linear Mixed Model



scaled to be weakly informative for fixed effects (Goodrich et al., 2020). JASP does not yet provide Bayes factors for linear mixed models, but this can be computed in R using the bayes_factor function in the `brms` package (Bürkner, 2017). JASP also provides variance and correlation estimates. The estimated standard deviation of the intercepts is 0.12, and the unexplained variance in work hours has a standard deviation of 0.88 (see Table 14).

Interpretation. Again, for guidance on what to report, see Depaoli and van de Schoot, 2017, Kruschke (2021), van Doorn et al., (2021), and van de Schoot et al. (2021). Note that JASP has some limitations when computing linear mixed models, because we cannot conduct some important components of Bayesian analysis, such as a posterior predictive check to examine the ability of the model to predict the observed data, or a sensitivity analysis to examine the influence of the prior on the posterior distribution. To report the results that we observed in JASP, we might say: "We conducted a linear mixed model with varying intercepts because we were interested in understanding how individual-level work demands (i.e., deviations from soldiers' respective group means) predict variance in soldier well-being, above and beyond independent group-level work demands.

We applied the default prior distribution in JASP software, which applies an exponential distribution on the standard deviation of the random effect and an independent normal distribution centered at 0 and scaled to be weakly informative for fixed effects. JASP automatically runs 2,000 burn-in samples, and 4,000 samples on three chains are included in the analysis with an adapt delta of .80 and a maximum tree depth of 10. The posterior distribution was adequately sampled, based on the effective sample size (ESS), and the MCMC chains converged, given that all R-hat values were less than 1.01 (see Table 13, see Appendix A for the mixing of chains, and see Appendix B autocorrelation plots). We observed a within-group effect of individual average work hours on well-being (β_1 = -0.05, 95% CrI [-0.06, -0.04]), such that for every one-hour increase, predicted well-being decreases by -0.05 points (see Table 13). We also observed a between-group effect of average group work hours on well-being (β_2 = -0.17, 95% CrI [-0.21, -0.14]), such that for every one-hour increase in total group hours, predicted well-being decreases by -0.17 points. Finally, the intercepts for within-group work hours have an estimated standard deviation of 0.12 (see Table 14)."

Bayesian – Varying Intercepts and Varying Slopes

Setup. We selected the same options as in the Bayesian varying intercepts model, but instead under "Random effects," we now check the boxes next to "Intercept," "W.HRS," and "G.HRS" to examine the effect of work hours and group hours on wellbeing.

Analysis. As before, JASP automatically populates tables similar to the varying intercept model. The coefficients (fixed effects) are highly comparable to the varying intercepts model (Table 15). JASP also provides model comparison estimates; Table 17 shows the leave-one-out (LOO) model comparison outputs. LOO is a type of cross-validation that assesses the robustness of models. It also relies on Pareto-smoothed importance sampling that helps identify points that have a bigger influence on the model. JASP also provides Watanabe-Akaike information criterion (WAIC), but the WAIC estimate for the varying intercepts and slopes model were unreliable in this case. JASP depends on the `loo` R package (Vehtari et al., 2023) to compute LOO and WAIC estimates.

Based on Table 17, we can see that the models perform the same for all practical purposes and so we favor the simpler, varying intercepts model. The estimated standard deviation of the intercept of group work demands is 0.12 and the standard deviation of the slope is 0.03. The correlation between the intercept and slope is -.24 and the unexplained variance of within group work hours has a standard deviation of 0.88 (see Table 16).

Interpretation. If we wanted to interpret the varying intercepts and slopes model, we could state: "We conducted a linear mixed model with varying intercepts and slopes to understand if individual level

Table 13. Predicting Soldier Well-Being:
Bayesian Fixed Effects Estimates for the Varying Intercepts Model

			95% CrI			
Term	Estimate	*SE*	Lower	Upper	ESS (bulk)	ESS (tail)
Intercept	4.74	.21	4.31	5.15	5704	4505
W.HRS	-0.05	.01	-.06	-0.04	9869	4027
G.HRS	-0.17	.02	-0.21	-0.14	5758	4531

Note. Total *N* = 7,382. W.HRS = group-mean centered hours (level one variable). G.HRS = group hours (level two variable). All terms had R-hat values equal to one suggesting convergence.

Table 14. Predicting Soldier Well-Being:
Bayesian Random Effects Estimates for the Varying Intercepts + Slopes Model

Level	Estimate
Level 2 (between)	
Intercept	0.12
Level 1 (within)	
Residual	0.88

Note. Total N = 7,382. Estimates are standard deviations.

work demands (i.e., deviations from the respective group average) predict soldier wellbeing above and beyond group-level work demands. We applied the default prior distributions in JASP software: i.e., an exponential distribution on the standard deviation of the random effect, and an independent normal distribution centered at 0 and scaled to be weakly informative for the fixed effects. JASP automatically runs 2,000 burn-in samples and then 4,000 samples on 3 chains with an adapt delta of .80 and a maximum tree depth of 10 that are included in the final analysis. The posterior distribution was adequately sampled, based on the effective sample size (ESS), and the model converged, where all R-hat values were less than 1.01 (see Table 15, see Appendix A for the mixing of chains, and see Appendix B for autocorrelation plots). We observed a within-group effect of individual average work hours on well-being (β_1 = -0.05, 95% CrI [-0.06, -0.04]), such that for every one-hour increase, predicted well-being decreases by -0.05 points (see Table 15). We also observed a between-group effect of average group work hours on well-being (β_2 = -0.17, 95% CrI [-0.21, -0.14]), such that for every one-hour increase in total group hours, well-being decreases by -0.17 points. The random effect of within-person work hours is small (SD = 0.03) which suggests that there is little variation between soldiers (see Table 16). Based on leave-one-out model comparison, this varying intercepts and slopes model does not have better model performance than the varying intercepts only model (see Table 17). Thus, we favor the simpler, varying intercepts only model and conclude that individual work demands do not sufficiently vary based on group level work demands."

Comparing and Contrasting Frequentist and Bayesian Results

In summary, both sets of analyses have a similar set-up and yield similar conclusions because we used Bayesian software default priors and provide two examples with sizable samples, but the results still have different technical interpretations. In the linear regression example, the R^2 values, intercept, and slope estimate are highly consistent in both the frequentist and Bayesian analyses. The important difference lies in how we can interpret the results. In the frequentist regression, we can reject the null hypothesis, that the intercept and slope estimates are equal to zero. In the Bayesian framework, the evidence provided by the data is in extreme favor of the alternative model (i.e., that the intercept and slope estimates are not equal to zero) relative to the null model. The general continuity between paradigm results is expected; Wetzels et al. (2011) reviewed 252 articles and found that generally, low p values corresponded to high BFs and high p values corresponded to low BFs – however, 70% of p values between .01 and .05 corresponded to "anecdotal evidence" under the Bayesian framework, suggesting that BFs provide more conservative estimates.

In the mixed linear model examples, we achieved highly comparable results; the intercept and main effects of individual hours and group hours were statistically significant in the frequentist analysis (which indicates that the 95% CIs did not contain 0), and had 95% CrIs that did not contain zero in the Bayesian analysis. In fact, the point estimates within the analysis. In fact, the point estimates within the frequentist framework are the same as the mean of the posterior

Table 15. Predicting Soldier Well-Being:
Bayesian Fixed Effects Estimates for the Varying Intercepts and Slopes Model

			95% CrI			
Term	Estimate	SE	Lower	Upper	ESS (bulk)	ESS (tail)
Intercept	4.73	.21	4.31	5.13	3078	3613
W.HRS	-0.05	.01	-0.06	-0.04	5040	4269
G.HRS	-0.17	.02	-0.21	-0.14	3115	4029

Note. Total N = 7,382. W.HRS = group-mean centered hours (level one variable). G.HRS = group hours (level two variable). All terms had R-hat values equal to one suggesting convergence of the MCMC chains.

Table 16. Predicting Soldier Well-Being:
Bayesian Random Effects Estimates for the Varying Intercepts + Slopes Model

Level	Estimate	Intercept	W.HRS
Level 2 (between)			
Intercept	0.12	–	
W.HRS	0.03	-.24	–
Level 1 (within)			
Residual	0.88		

Note. Total N = 7,382. Estimates are standard deviations, and a correlation in the square matrix.

Table 17. Leave One Out (LOO) - Model Comparison

Model	LOO	SE (LOO)
Varying Intercepts	19174.5	117.0
Varying Intercept + Slopes	19175.0	117.0

Note. LOO values and standard error for the varying intercept and varying intercept + slopes model.

distributions in the Bayesian framework. The variance and correlation estimates were also nearly identical between statistical frameworks. Again, the key difference lies in how we can interpret the results. For instance, in the frequentist example, focusing on the fixed effect slope of group work demands on well-being, if we were merely to reject the null hypothesis, we would simply state that the slope coefficient of group work demands is reliably different from zero. However, we usually also focus on the estimated value and the associated 95% CI, which together tells us that the estimate is the best estimate of the population value, and given uncertainty in the estimate there is a 95% chance that the CI contains the parameter of interest. Comparably, in the Bayesian paradigm, the 95% CrI of group hours did not contain zero, and it tells us – in a much more straightforward manner – that 95% of the most probable parameters are contained in the interval.

The model comparison indices also differ between paradigms. Whereas AIC (based on Kullback–Leibler divergence) is a common index of model comparison the frequentist framework and evaluates model fit, the Bayesian LOO (a type of cross validation) is common in the Bayesian framework and examines the model's out of sample prediction. Here, the frequentist AIC favors the more complex varying intercepts and slopes model, whereas the Bayesian LOO favors the simpler varying intercepts and fixed slope model. However, it is worth noting that the frequentist AIC difference value was 2, which is the smallest AIC value that suggests a meaningful difference between models (Burnham & Anderson, 2004), and the Bayesian LOO suggested that neither model outperformed the other, which led us to favor the simpler model. When considering these factors, the broader conclusions are not substantially different. As we said in the introduction, these statistical frameworks have different philosophical underpinnings and different methods of model comparison, which yield slightly different model comparison conclusions in this specific case, although the estimates for the fixed effects and random effects are nearly identical.

Conclusion

For all disciplines, including our target audience of organizational researchers and graduate students, JASP provides a user-friendly platform to conduct basic Bayesian analyses and communicate Bayesian results without having to do any coding. JASP also helps researchers run Bayesian analysis by virtue of its thoughtful selection of default settings for the most common analyses conducted in organizational research. JASP then serves as a bridge to using R. More specifically, as researchers become more adept in their Bayesian knowledge and modeling abilities, there are great advantages to using R packages, such as `brms` and `rstanarm`, to build Bayesian models (Bürkner, 2017; Goodrich et al., 2020).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716-723. https://doi.org/10.1109/TAC.1974.1100705

Best, N. G., Spiegelhalter, D. J., Thomas, A., & Brayne, C. E. G. (1996). Bayesian analysis of realistically complex models. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *159*(2), 323-342. https://doi.org/10.2307/2983178

Bliese, P. D., & Halverson, R. R. (1996). Individual and nomothetic models of job stress: An examination of work hours, cohesion, and well-being. *Journal of Applied Social Psychology*, *26*(13), 1171–1189. https://doi.org/10.1111/j.1559-1816.1996.tb02291.x

Bliese, P. (2022). *Multilevel modeling in R (2.7)–A brief introduction to R, the multilevel package and the nlme package.*

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*(2), 261–304. https://doi.org/10.1177/0049124104268644

Bürkner P (2017). `brms`: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Clyde, M. A., Ghosh, J., & Littman, M. L. (2011). Bayesian adaptive sampling for variable selection and model averaging. *Journal of Computational and Graphical Statistics*, *20*(1), 80–101. https://doi.org/10.1198/jcgs.2010.09049

Clyde, M. A. (2018). *BAS: Bayesian variable selection and model averaging using Bayesian adaptive sampling.* (Version 1.5.3)[Computer software].

Consonni, G., Fouskakis, D., Liseo, B., & Ntzoufras, I. (2018). Prior distributions for objective Bayesian analysis. *Bayesian Analysis*, *13*, 627-679. https://doi.org/10.1214/18-BA1103

Depaoli, S., & van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist. *Psychological Methods*, *22*, 240–261. https://doi.org/10.1037/met0000065

Depaoli, S., Winter, S. D., & Visser, M. (2020). The importance of prior sensitivity analysis in Bayesian statistics: Demonstrations using an interactive shiny app. *Frontiers in Psychology*, *11*, 1-18. https://doi.org/10.3389/fpsyg.2020.608045

Faulkenberry, T. J., Ly, A., & Wagenmakers, E.-J. (2020). Bayesian inference in numerical cognition: A tutorial using JASP. *Journal of Numerical Cognition*, *6*(2), 231-259. https://doi.org/10.5964/jnc.v6i2.288

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://www.jstor.org/stable/2246093

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2015). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b16018

Goodrich B., Gabry J., Ali I., Brilleman S. (2020). *rstanarm: Bayesian applied regression modeling via Stan*. (Version 2.21.1)[Computer software].

Gronau, Q. F., Ly, A., & Wagenmakers, E.-J. (2020). Informed Bayesian *t*-tests. *The American Statistician*, *74*(2), 137–143. https://doi.org/10.1080/00031305.2018.1562983

The jamovi project (2023). *jamovi* (Version 2.3) [Computer Software].

JASP Team (2023). *JASP* (Version 0.17) [Computer software].

JASP (2023). *Release notes. JASP - Free and user-friendly statistical software*. https://jasp-stats.org/release-notes/

Jebb, A. T., & Woo, S. E. (2015). A Bayesian primer for the organizational sciences: The "two sources" and an introduction to BugsXLA. *Organizational Research Methods*, *18*(1), 92-132. https://doi.org/10.1177/1094428114553060

Jeffreys, H. (1961). *Theory of probability* (3rd ed.), Clarendon Press.

Kelter, R. (2020). Bayesian alternatives to null hypothesis significance testing in biomedical research: A non-technical introduction to Bayesian inference with JASP. *BMC Medical Research Methodology*, *20*. https://doi.org/10.1186/s12874-020-00980-6

Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods*, *15*(4), 722-752. https://doi.org/10.1177/1094428112457829

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.* Academic Press.

Kruschke, J. K. (2021). Bayesian analysis reporting guidelines. *Nature Human Behaviour*, *5*(10), 1282–1291. https://doi.org/10.1038/s41562-021-01177-7

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course.* Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

Liang, F., Paulo, R., Molina, G., Clyde, M. A., & Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, *103*, 410-423. https://doi.org/10.1198/016214507000001337

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J. N., Morey, R. D., & Wagenmakers, E.-J. (2019). JASP: Graphical statistical software for common statistical designs. *Journal of Statistical Software*, *88*, 1–17. https://doi.org/10.18637/jss.v088.i02

Marsman, M., & Wagenmakers, E.-J. (2017). Bayesian benefits with JASP. *European Journal of Developmental Psychology*, *14*(5), 545–555. https://doi.org/10.1080/17405629.2016.1259614

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and STAN* (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9780429029608

Quintana, D. S., & Williams, D. R. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using JASP. *BMC Psychiatry*, 18. https://doi.org/10.1186/s12888-018-1761-4

R Core Team (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Rouder, J. N., & Morey, R. D. (2012). Default Bayes factors for model selection in regression. *Multivariate Behavioral Research*, *47*(6), 877–903. https://doi.org/10.1080/00273171.2012.734737

Scott, J. G., & Berger, J. O. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics*, *38*(5), 2587–2619. https://doi.org/10.1214/10-AOS792

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 461-464. http://www.jstor.org/stable/2958889.

Stan Development Team. (2022a). Stan reference manual (Version 2.32). https://mc-stan.org/docs/2_32/reference-manual-2_32.pdf

Stan Development Team. (2022b). Runtime warnings and convergence problems. https://mc-stan.org/misc/warnings.html#maximum-treedepth

Stefan, A. M., Gronau, Q. F., Schönbrodt, F. D., & Wagenmakers, E.-J. (2019). A tutorial on Bayes Factor Design Analysis using an informed prior. *Behavior Research Methods*, *51*(3), 1042–1058. https://doi.org/10.3758/s13428-018-01189-8

van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods*, *22*(2), 217-239. http://dx.doi.org/10.1037/met0000100

van de Schoot, R., Sijbrandij, M., Depaoli, S., Winter, S. D., Olff, M., & Van Loey, N. E. (2018). Bayesian PTSD-trajectory analysis with informed priors based on a systematic literature search and expert elicitation. *Multivariate Behavioral Research*, *53*(2), 267–291. https://doi.org/10.1080/00273171.2017.1412293

van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., & Yau, C. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, *1*(1), Article 1. https://doi.org/10.1038/s43586-020-00001-2

van den Bergh, D., Doorn, J. van, Marsman, M., Draws, T., van Kesteren, E.-J., Derks, K., Dablander, F., Gronau, Q. F., Kucharský, Š., Raj, A., Sarafoglou, A., Voelkel, J. G., Stefan, A. M., Ly, A., Hinne, M., Matzke, D., & Wagenmakers, E.-J. (2019). *A tutorial on conducting and interpreting a Bayesian ANOVA in JASP*. PsyArXiv. https://doi.org/10.31234/osf.io/spreb

van den Bergh, D., Clyde, M. A., Gupta, A. R., de Jong, T., Gronau, Q. F., Marsman, M., ... & Wagenmakers, E. J. (2021). A tutorial on Bayesian multi-model linear regression with BAS and JASP. *Behavior Research Methods*, *53*, 2351–2371. https://doi.org/10.3758/s13428-021-01552-2

van Doorn, J., van den Bergh, D., Boehm, U., Dablander, F., Derks, K., Draws, T., Evans, N. J., Gronau, Q. F., Hinne, M., Kucharsky, S., Ly, A., Marsman, M., Matzke, D., Komarlu Narendra Gupta, A. R., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E.-J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, *28*, 813-826. https://doi.org/10.3758/s13423-020-01798-5

van Erp, S., Mulder, J., & Oberski, D. L. (2017). Prior sensitivity analysis in default Bayesian structural equation modeling. *Psychological Methods*, *23*(2), 363-388. https://doi.org/10.1037/met0000162

Vehtari, A., Gabry J., Magnusson, M., Yao, Y., Bürkner, P., Paananen, T., & Gelman, A. (2023). loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models. R package version 2.6.0, https://mc-stan.org/loo/.

Vrieze, S. I. (2012). Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, *17*(2), 228-234. https://doi.org/10.1037/a0027127

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., … Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58-76. https://doi.org/10.3758/s13423-017-1323-7

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 *t* tests. *Perspectives on Psychological Science*, *6*(3), 291-298. https://doi.org/10.1177/1745691611406923

Wetzels, R., Raaijmakers, J. G., Jakab, E., & Wagenmakers, E. J. (2009). How to quantify support for and against the null hypothesis: A flexible WinBUGS implementation of a default Bayesian *t* test. *Psychonomic Bulletin & Review*, *16*, 752-760. https://doi.org/10.3758/PBR.16.4.752

Zitzmann, S., & Hecht, M. (2019). Going beyond convergence in Bayesian estimation: Whyprecision matters too and how to assess it. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(4), 646–661. https://doi.org/10.1080/10705511.2018.1545232

Zyphur, M. J., & Oswald, F. L. (2015). Bayesian estimation and inference: A user's guide. *Journal of Management*, *41*(2), 390-420. https://doi.org/10.1177/0149206313501200

Citation:
Courey, K. A., Oswald, F. L., & Culpepper, S. A. (2024). Clearer analysis, interpretation, and communication in organizational research: A Bayesian guide. *Practical Assessment, Research, & Evaluation*, 29(1). Available online: https://doi.org/10.7275/pare.1975

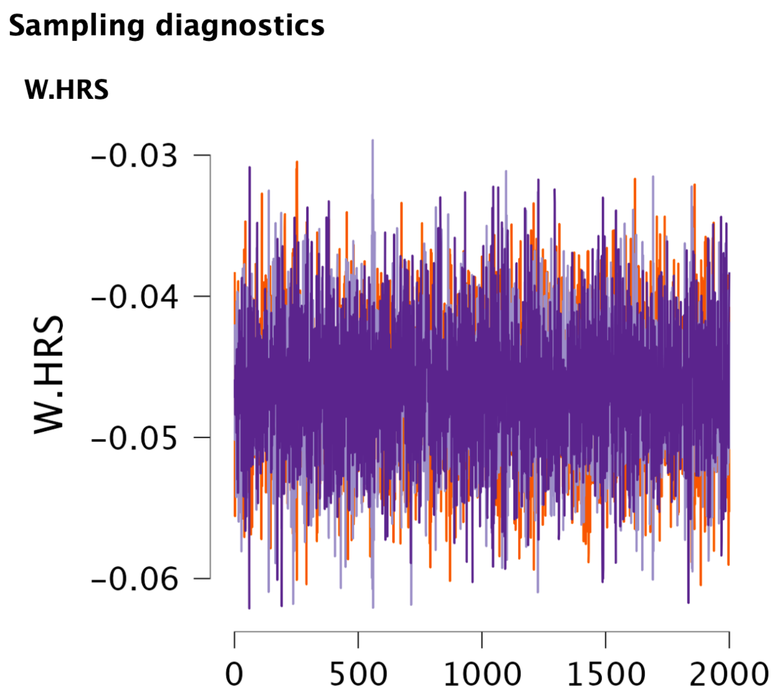
Corresponding Author:
Karyssa A. Courey
Rice University

Email: kac23 [at] rice.edu

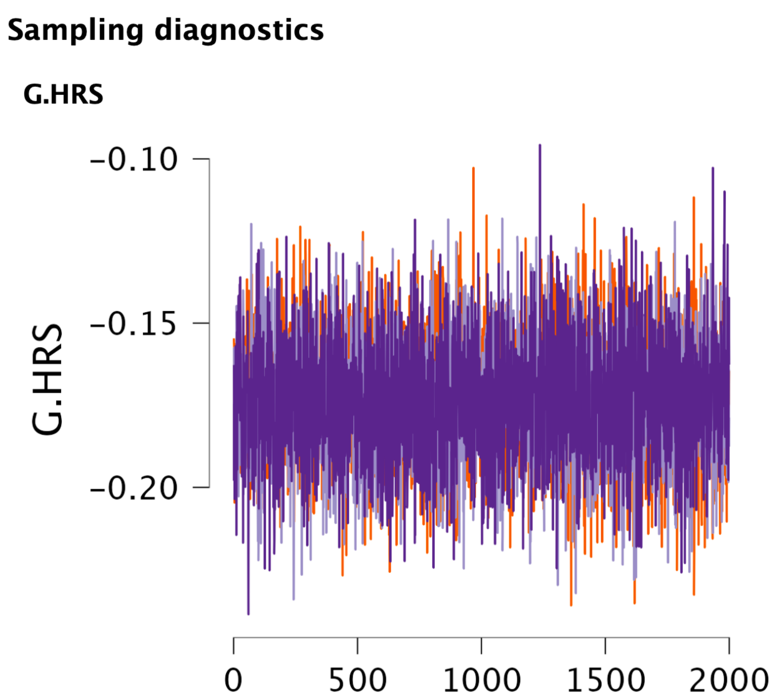
Appendix A

MCMC Three Chain Mixing: Varying Intercepts Model

Individual Hours (Level 1 Variable)

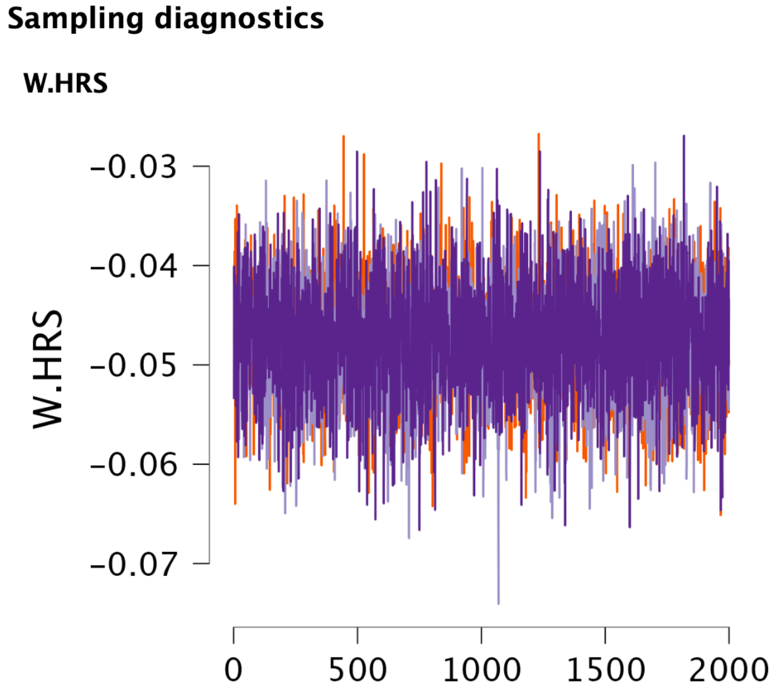


Group Hours (Level 2 Variable)

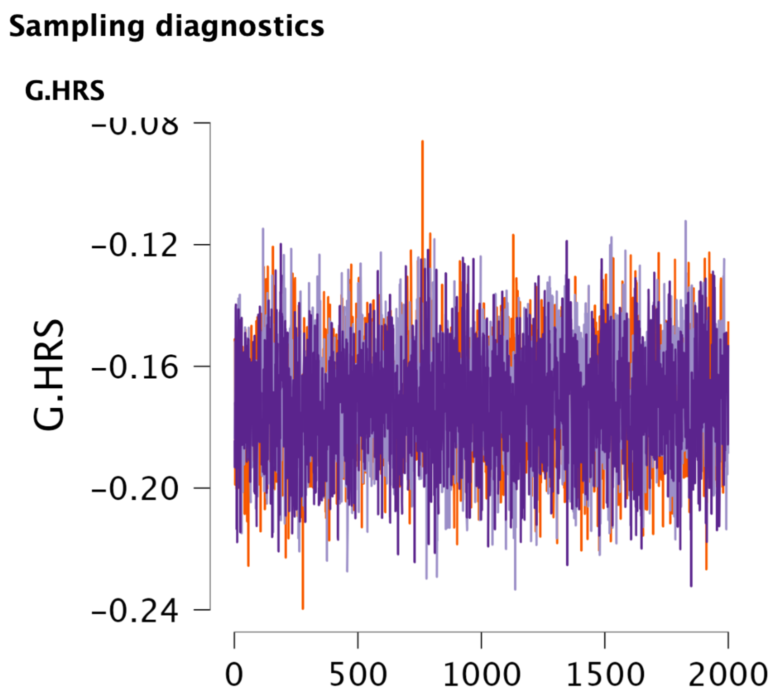


MCMC Three Chain Mixing: Varying Intercepts Model

Individual Hours (Level 1 Variable)



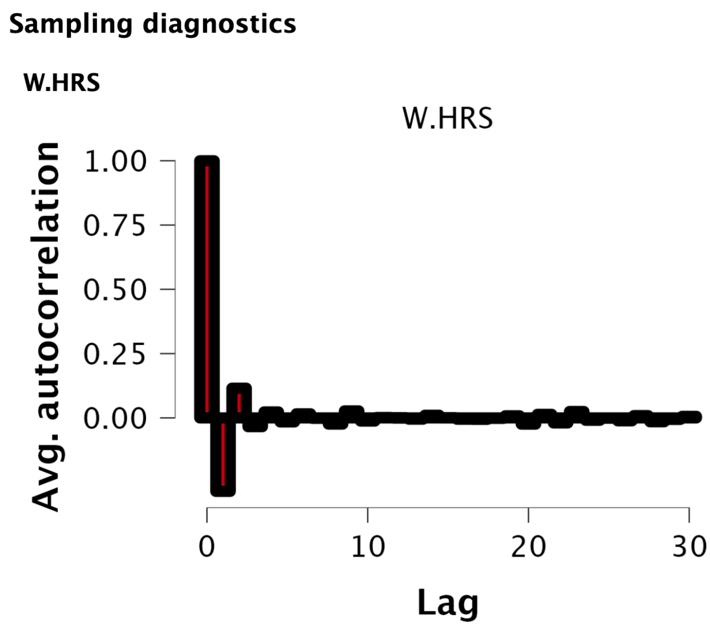
Group Hours (Level 2 Variable)



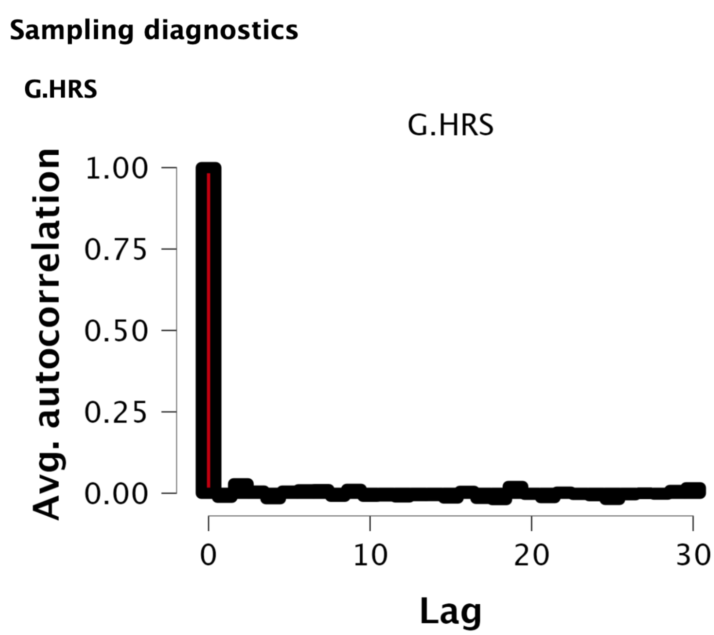
Appendix B

MCMC Three Chain Autocorrelations: Varying Intercepts Model

Individual Hours (Level 1 Variable)

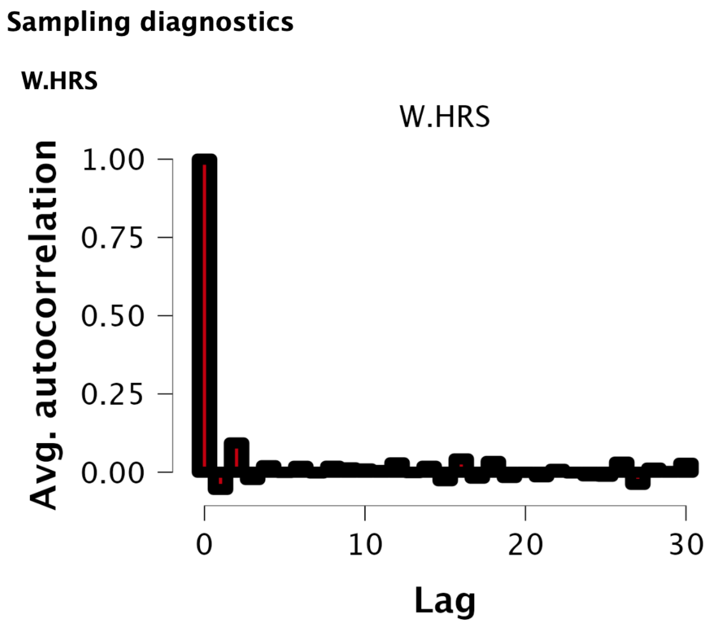


Group Hours (Level 2 Variable)



MCMC Three Chain Autocorrelations: Varying Intercepts + Slopes Model

Individual Hours (Level 1 Variable)



Group Hours (Level 2 Variable)