From clinical trials to general U.S. population: efficacy of self-medicating with psychedelics to improve management of a psychiatric condition

Bayesian Statistics

Maaike Walraad (7079540)

1. Introduction

Psychedelics (serotonergic hallucinogens) have been used for thousands of years around the world for their psychoactive effects on perception, mood, and cognition. In 2013, the estimated number of psychedelics users in the USA alone was 30 million (Krebs & Johansen, 2013). A growing number of individuals appear to be taking psychedelics under supervision, to treat mental health concerns such as depression and anxiety (The Medical Republic, 2021). In the last few years, there has been renewed interest in the use of psychedelics and a significant acceleration in research activity, leveraging their neuroplastic effect as a novel approach for the treatment of psychiatric disorders. Psychedelic mushrooms (PM) are a popular type of psychedelics in this research field.

This paper aims to answer the following research question: *"What is the relation between self-medicative PM use and severity of depression among persons with major depressive disorder?"*

2. Method

2.1 Data

In 2021, a large survey study was published that aimed to explore the prevalence of PM use and its association with various measures of (mental) health among the general U.S. population (Morlock, 2021). A specific subset was taken from the survey data, only including PM users during the last 12 months (n=257). The independent variable in this report was whether the usage of PM was for self-medication purposes (*PM*). The outcome variable was the Patient Health Questionnaire-9 (*PHQ-9*) score, which is a measure for the severity of depression. The inclusion criteria for this paper included: aged 18-65, experienced depression last year and at least severe depression (PHQ9 score > 15) when the survey was conducted. A PHQ9 score above 15 indicates the presence of major depressive disorder (Kroenke, Spitzer & Williams, 2001). This selection resulted in a sample size of 64, with 28 persons who used PM to self-medicate, and the 36 persons who did not.

2.2 Priors

As for the use of informative priors, 74 clinical studies registered on clinicaltrials.gov were analyzed. Eligible studies were selected based on whether they used the PHQ-9 to measure depression, a type of PM as treatment, and similar selection criteria. This selection resulted in one study (see Table 1).

Table 1. Completed clinical trial, including results.

Study	*Year*	*Sample size*	*pre-PHQ-9 M(SD)*	*post-PHQ-9 M(SD)*
Davis et al. (2021)	2021	24	16.5 (3.1) 17.9 (3.3)	4.8 (2.9) 18.8 (4.3)

2.3 Statistical model

To answer the question of whether there is a relationship between the use of self-medicative PM use among the general (U.S.) population and depression, simple linear regression analysis was performed. The regression model was defined as:

$$PHQ9_i = b_0 + b_1 * PM + e_i$$

with

$$e_i \sim N(0, \sigma^2)$$

The regression parameter estimates were obtained through a Gibbs sampler, a type of Markov chain Monte Carlo (MCMC) method. Rather than obtaining the joint posterior of the regression parameters *exactly* (based on Bayes theorem), the conditional posterior distributions are approximated. Samples are then drawn from the conditional posterior distributions to obtain the joint posterior.

Gibbs sampling approximated the posterior conditional distributions for the intercept (b_0), the regression parameter for PM use (b_1), and the residual variance (σ^2) through the following steps. The first step of the Gibbs algorithm is the specification of priors for each parameter. For b_0 and σ^2, semi-conjugate priors were specified. Semi-conjugate means that the prior distribution, with respect to the density, results in a conditional posterior distribution that has the same form as the prior. Though, one mostly just uses the word "conjugate" to denote said priors. Vague priors for the mean and variance of b_0 were chosen:

$$b_o \sim N(0, 10^3)$$

When setting the prior variance very large, the prior specification for b_0 does not influence the posterior distribution. For σ^2, priors based on the previous study by Davis et al. (2021) were specified. Though the sum of squared residuals (S_0) in the previous study is unknown, it was set to a reasonable value of 100. Setting the priors for the shape and scale hyperparameters resulted in:

$$\sigma^2 \sim IG\left(\frac{N_0}{2}, \frac{S_0}{2}\right) = IG\left(\frac{24}{2}, \frac{100}{2}\right)$$

The prior for b_1 is a non-conjugate prior, and is specified based on the historical data from the clinical trial by Davis et al. (2021):

$$b_1 \sim lst(-11.7, 3^2, 21)$$

The mean (-11.7) and variance (3^2) are based on table 1. The prior degrees of freedom (21) are calculated by subtracting the number of estimated parameters by the number of participants in the study by Davis et al. (2021).

The second step is determining the full conditional distributions. To derive the conditional posterior distributions for each regression parameter θ, only the numerator in Bayes theorem is used:

$$p(\theta|y) \propto f(y|\theta)p(\theta)$$

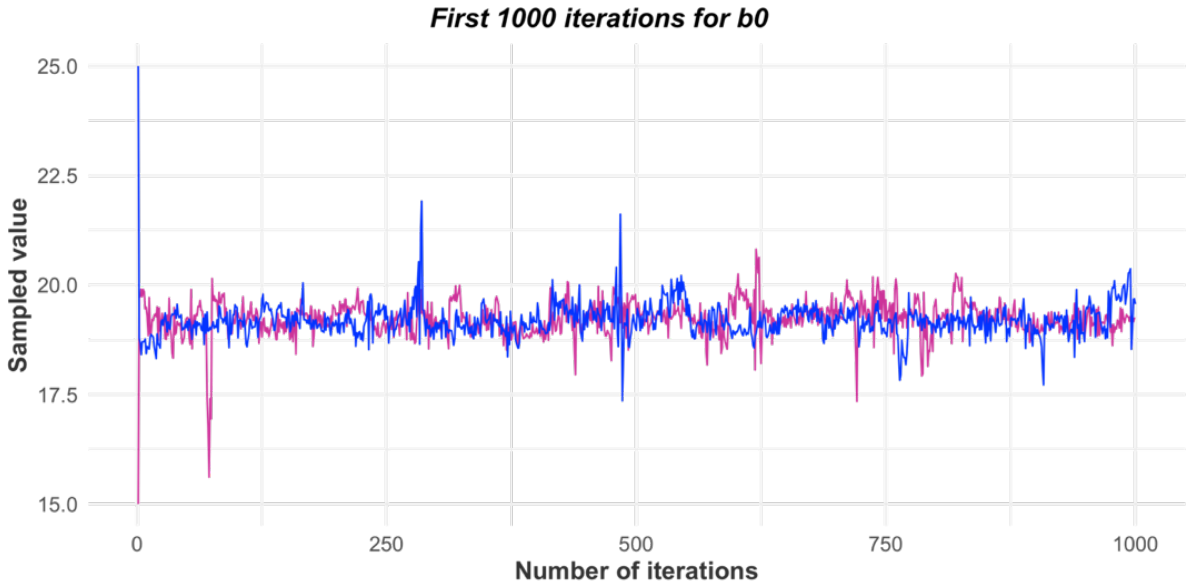
where $p(y|\theta)$ equals the density of the data, and $p(\theta)$ is the prior. Our posterior is then formally not a probability distribution, but it will have the exact same shape. All the proportions between the different probabilities are still the same, so calculating the mean, quantile interval, and variance still gives us the most likely parameters given our data. The multiplication of density x prior is simplified by removing any parts that don't contain θ and multiplying out the independence. In the appended

file ‘code.R’, see lines 87-89 and 101-104 for the (simplified) estimation of the posterior distribution for b_0 and σ^2, respectively.

For b_1, the conditional posterior distribution is approximated using a Metropolis Hastings (MH) algorithm. With an MH algorithm, a function proportional to a conditional posterior is derived, to then use this function to sample from a proposal distribution. This algorithm is often used in the case of non-conjugate priors. The prior for b_1 follows a t-distribution, as this distribution expresses more uncertainty. This prior distribution was chosen in this study because the prior parameters were weakly informed, given they were based on a single clinical trial. The proposal distribution was carefully chosen based on convergence diagnostics, as it should not be too wide or too narrow. The proposal was a normal distribution with a mean of 0 and a standard deviation of 3. The proposal did not depend on the most recently accepted value, which means it is an independent MH sampler. See lines 69-74 and 92-98 in ‘code.R’ for the MH estimation of b_1.

Per parameter, two Markov chains with different starting values were recorded. To prove that the Gibbs algorithm works, the first 1000 iterations of the two Markov chains for b_0 are shown in Figure 1 below. The chains had different starting values, but quickly started mixing (after 50 its.). This is achieved in Gibbs sampling, because the long-term probability of taking on each value is independent of the initial value (Gundersen, 2022).

Figure 1. First 1000 iterations for b_0 using different starting values for the two chains (15 and 25).



In total, 26000 Gibbs samples were drawn, of which the first 1000 were warm up and removed for inference. After obtaining the conditional posterior distributions for all parameters, the mean or Expected A Posteriori (EAP), standard deviation (SD) and central credible interval (CCI) were calculated for each parameter. To assess convergence of the Gibbs algorithm, trace plots and autocorrelation were checked. Additionally, the acceptance rate was calculated to evaluate the convergence of the Metropolis-Hastings algorithm.

2.4 Frequentist vs. Bayesian

In this paper, the different approaches between Frequentist and Bayesian statistics were considered to answer the research question. A standard least squares regression model was compared to a Gibbs sampled regression model with informative priors. The primary motive for comparing a Frequentist to a Bayesian approach is the inclusion of priors in the latter. The goal here is to examine the effect of prior knowledge on the results. Though, the research question of this study will only be answered in the context of Bayesian statistics.

3. Results

3.1 Regression estimates

The parameter estimates are given in Table 2, based on the first Markov chain. The posterior mean for the intercept parameter is 19.16. The EAP for *PM* is 1.70 (SD=0.59), with a 95% CCI of 0.26 – 2.76. The mean of the residual error variance is 4.47, with a SD of 16.17.

Table 2. Gibbs regression parameter estimates, including mean, standard deviation, and 95% CCI.

	Mean	SD	2.5%	97.5%
Intercept (b_0)	19.16	0.36	18.49	19.90
PM (b_1)	1.70	0.59	0.26	2.76
Variance (σ^2)	4.47	16.17	0.98	21.01

3.2 Diagnostics

Figure 2 shows traceplots of the sampled parameter values over the 25000 iterations (excluding warm up). The desired pattern in the distribution is one that is stationary, as it gives an indication of convergence. For b_0 and b_1, the chains have mixed properly, as they have approximately the same mean and variance. Simultaneously, the traceplots show no extreme values that were sampled. For the regression coefficient that has been sampled using a Metropolis Hastings step, b_1 , the acceptance rate was 15.8%. The acceptance rate given an indication of the quality of the proposal distribution: for an independent MH sampler, a higher rate is generally better. The traceplot for σ^2 shows extremely small variance, with some large deviations. This corresponds to the large standard deviation and CCI of σ^2 shown in table 2.

Figure 3 shows autocorrelation for the two chains for the estimated parameters. Autocorrelation denotes the correlation of Gibbs samples of a parameter with itself lagged in time. For b_0 and σ^2, there is little autocorrelation (<0.50). For b_1, the autocorrelation starts out higher, but decreases below 0.50 within 10 lags.

For the parameters b_0 and b_1, it is acceptable to assume that convergence has been reached. For σ^2, the diagnostics show signs of not converging, particularly due to the extreme outlier values beyond the 20000st iteration, indicating the need for more samples. No additional samples were drawn, as 25000 is already many samples. Therefore, the results should be interpreted with this uncertainty in mind.

3.3 Frequentist vs. Bayesian

Table 3 shows the regression estimates of the frequentist regression analysis. The regression coefficient for *PM* is 1.82, with a 95%-confidence interval that contains 0. The residual standard error for the frequentist linear regression is 14.41 (df=62).

Table 3. Frequentist linear regression analysis.

	Estimate	SD	95%-confidence interval
Intercept (β_0)	19.11	0.63	17.85 – 20.38
PM (β_1)	1.82	0.96	-0.10 – 3.73

Figure 2. *Traceplots for b_0, b_1 and σ^2.*

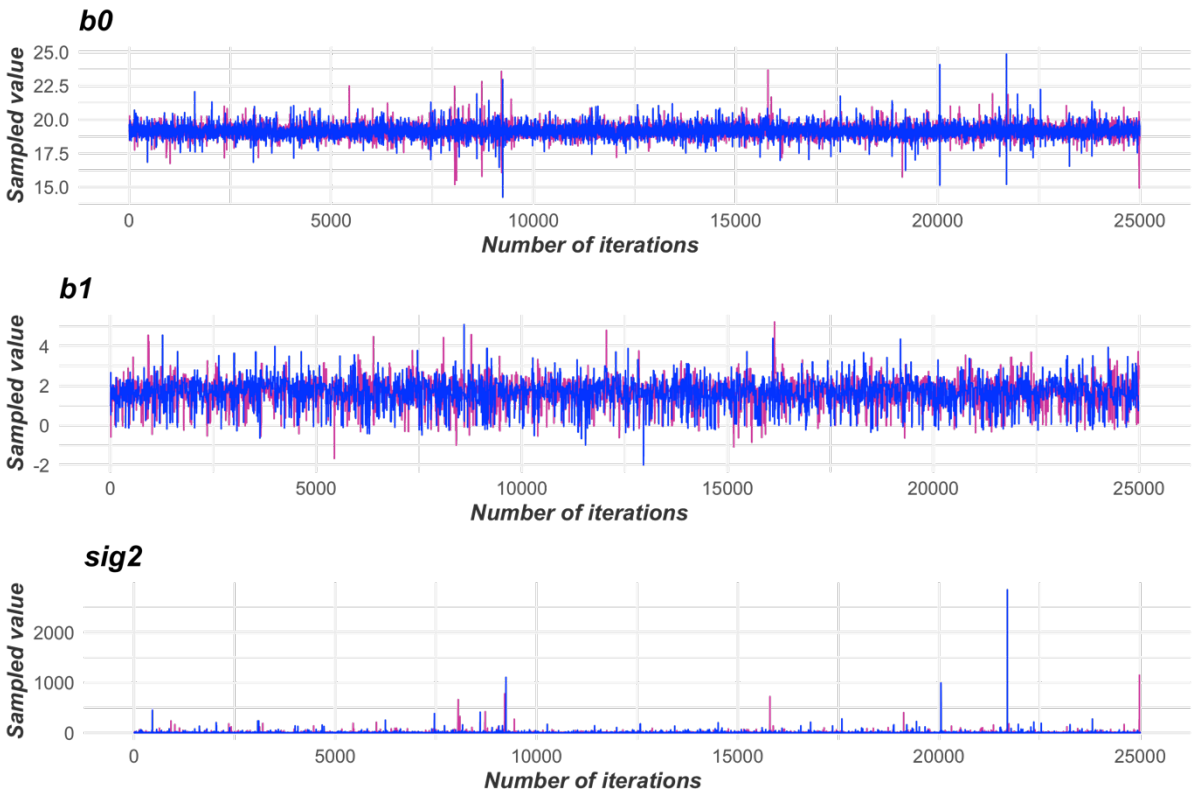
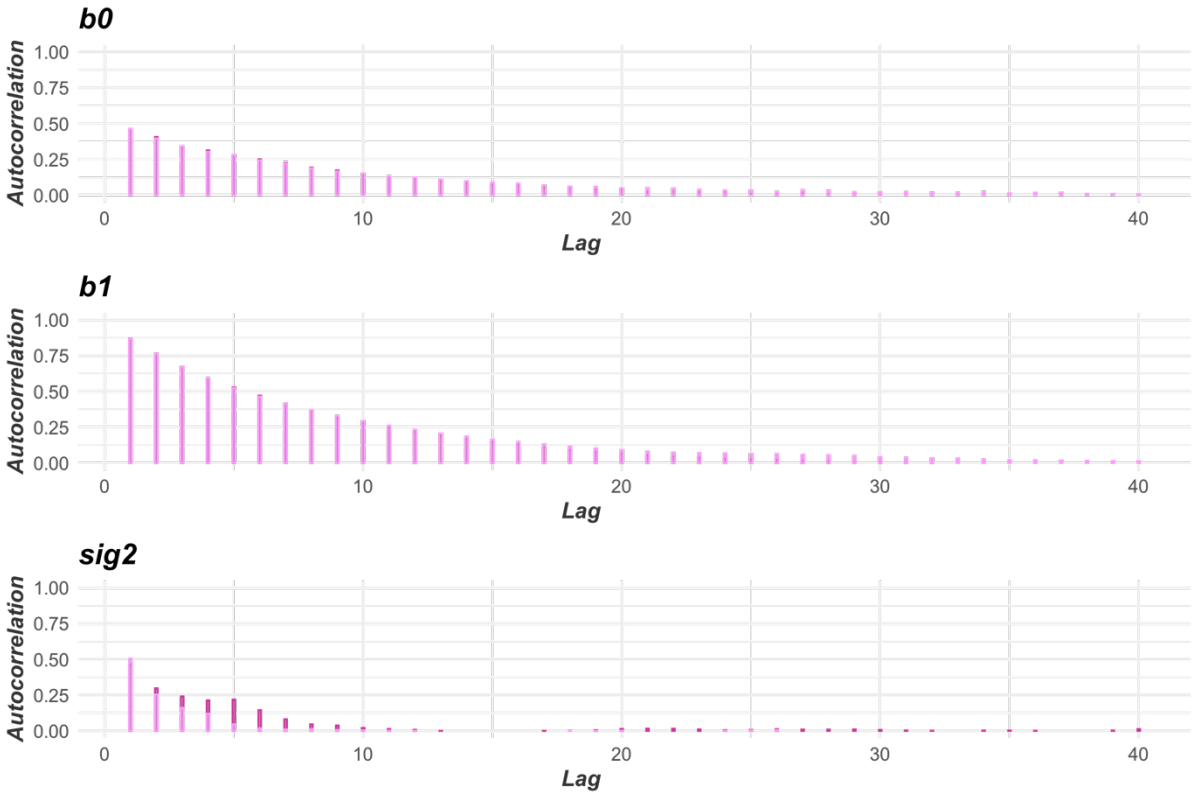


Figure 3. *Autocorrelation plots for b_0, b_1 and σ^2.*



4. Discussion

This study aimed to assess the relation between self-medicative PM use and severity of depression among persons with major depressive disorder. The 95% CCI for the predictor PM did not contain the value zero, which means one can be 95% certain that there is a positive relation between PM usage and severity of depression. This result is opposite to the prior knowledge.

An important philosophical difference between Bayesian and frequentist in this study is the use of priors. Often, one already has prerequisite knowledge or ideas about the direction or magnitude of a statistic. In Bayesian statistics, the use of prior knowledge allows for more usage of previous obtained scientific knowledge. Data can be wrong, but so can prior knowledge: conflicting hypotheses and data can be very useful. If used correctly, prior information enables faster convergence to the truth. Priors can be subjective (personal beliefs) or based on historical data. The choice to, and in which fashion, incorporate prior knowledge should be subjected to a few considerations. In this study, the priors were based on historical data. An important question may be: "How much weight should be given to the historical data?". For this study, there was only one applicable historical study, which was therefore weighted for 100% in the prior. A potential downside to this may be that a single study can yield extreme results (when no priors are used!), which would be cancelled out in the case of using multiple studies to construct a prior.

From a statistical viewpoint, prior information is equivalent to additional data, and data is cumulative. This means that there will always be more certainty around an estimate, as the variance of the posterior distribution shrinks. This can be seen from the results, as the standard deviations and credible intervals of the regression coefficients were smallest for the Gibbs regression model. This finding proves that the frequentist and Bayesian approaches yield different results in the context of this study. Interpretation wise, one can only deduce statistical inference with certainty from the Bayesian analysis: not from the frequentist analysis. Measuring (mental) health outcomes in a randomized controlled (RCT) setting, though necessary to obtain causal inference, is a costly procedure compared to a cross-sectional survey for example. Therefore, valorizing valuable knowledge from RCT studies by incorporating it into subsequent studies seems justified. The limitation of the current study design, however, is that the data are collected in a cross-sectional manner, meaning one cannot infer any causal relations from the results. Using historical data from a different study design may therefore be unjustified and ultimately ineffective.

5. References

Gundersen, G. (2022). Gibbs Sampling Is a Special Case of Metropolis-Hastings. [Online; accessed 20. Jun. 2023]. URL: https://gregorygundersen.com/blog/2020/02/23/gibbs-sampling

Krebs, T. S., & Johansen, P. Ø. (2013). Over 30 million psychedelic users in the United States. *F1000Research*, *2*, 98. https://doi.org/10.12688/f1000research.2-98.v1

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Morlock, R. (2021). Data from: Psychedelic mushrooms in the USA: knowledge, patterns of use, and association with health outcomes, Dryad, Dataset, https://doi.org/10.5061/dryad.bzkh189b6

The Medical Republic (2021). Self-medicating with psychedelics on the rise. [Online; accessed 20. Jun. 2023]. URL: https://www.medicalrepublic.com.au/self-medicating-with-psychedelics-on-the-rise/41414