American Journal of Physical Medicine & Rehabilitation. 2019; 98(6): 516-520

Bayesian methods in the field of rehabilitation

Quintela-del-Río, Alejandro PhD; Rodríguez-Romero, Beatriz PhD; Robles-García, Verónica PhD; Arias-Rodríguez, Pablo PhD; Cudeiro-Mazaira, Javier MD; Martínez-Rodríguez, Alicia PhD

From the Department of Mathematics, Universidade da Coruña, A Coruña, Spain (AQ-d-R); Department of Physiotherapy, Medicine and Biomedical Sciences, Universidade da Coruña, A Coruña, Spain (BR-R, VR-G, JC-M, AM-R); and Department of Physical and Sports Education, Universidade da Coruña, A Coruña, Spain (PA-R).

Abstract

Bayesian techniques, as an alternative method of statistical analysis in rehabilitation studies, have some advantages such as handling small sample sizes, allowing incorporation of previous experience of the researchers or clinicians, being suitable for different kinds of studies, and managing highly complex models. These characteristics are important in rehabilitation research. In the present article, the Bayesian approach is displayed through three examples in previously analyzed data with traditional or frequentist methods. The studies used as examples have small sample sizes and show that the Bayesian procedures enhance the statistical information of the results. The Bayesian credibility interval includes the true value of the corresponding parameter diminishing uncertainty about the treatment effect. In addition, the Bayes factor value quantifies the evidence provided by the data in favor of the alternative hypothesis as opposed to the null hypothesis. Bayesian inference could be an interesting and adaptable alternative statistical method for physical medicine and rehabilitation applications.

Key Words:

Bayes Theorem, Rehabilitation, Sample Size, Statistics

Several conclusions from clinical studies in the rehabilitation field are derived from the interpretation of the outcomes from statistical analysis of the obtained data. Such studies are usually carried out through inferences about the parameters of the corresponding variables, mainly calculating confidence intervals or conducting hypothesis testing (difference of means test, analysis of variance, regression analysis, etc.). In most cases, results and conclusions are based on P values.[1] However, common misconceptions about confidence intervals and P values exist.[2,3] For these and other reasons, Bayesian statistics have been applied as a counterpart to frequentist or traditional approaches to statistical analyses.[4,5] The Bayesian approach is based on a complete set of statistical methods to gain the benefit of the numerical results of an experiment, for example, by means of credibility intervals (CI) and the Bayes factor (BF), and it can provide considerable support to the statistical conclusions.[6,7] Moreover, Bayesian analysis can increase the statistical power to acceptable levels,[8] which is not always achieved in frequentist studies because of the difficulty of obtaining large samples.[9]

The aims of this article are to provide an insight into the basics of Bayesian statistics and to encourage a pursuit of a deeper understanding of how they can be applied to physical medicine and rehabilitation research. This brief introduction is accompanied by examples of clinical data previously published in studies about musculoskeletal pain and a new therapeutic strategy for people with Parkinson disease.

METHODS AND CONTENTS

The Bayesian approach in the evaluation of experimental results has been proposed in the field of medicine[6,10] and has become more common in many fields such as psychology[11,12] and technical disciplines,[13,14] whereas in other areas, such as nursing[15] or physical therapy and rehabilitation,[16] it is less common. For example, when key words such as “Bayes,” “Bayesian,” or similar are used in searches in the rehabilitation category of the Sciences Citation Index, very few articles are obtained. As far as we know, only the recent article of Nuzzo[13] is directly related to Bayesian inference in the setting of the linear correlation coefficient estimation.

Bayesian analysis works under the general assumption that the credibility of a theoretical model for describing experimental data can be combined from both objective and subjective information. Objective information is based only in the probability provided by the data (likelihood). The subjective knowledge or information (prior probability) comes from the clinician's own experience and competence, and it is quantified by a prior distribution (prior). Bayesian statistics combine both factors in the named posterior probability (posterior).

For example, to test the effect of the same therapy on a different target population to the one previously studied, the Bayesian method allows evidence generated from previous studies to be taken into account. In this way, inferences about parameters, models, or hypotheses can be updated as evidence accumulates, and the posterior probabilities can be used as prior probabilities in a new step of the experiment or clinical trial. However, if there is no previous experience to draw from, or if the researcher does not have (or does not wish to express) any initial information by means of a prior probability, Bayesian analysis can use the named uninformative prior probabilities (which consists of assigning the same probability for the possible results of the experiment). This case agrees with the frequentist statistical way of thinking,[17] and the corresponding results can be a valuable complement to standard studies and decision-making processes.

Bayesian analysis is particularly useful in the rehabilitation field for two reasons. The first benefit relates to sample sizes.[18] In this area, it is common to find small samples in studies conducted.[9] However, in Bayesian analysis, no minimum sample size is required for calculations because they do not rely on asymptotic results.[12] The second reason is that many usual assumptions, such as the normality of the variables, do not need be fulfilled. This is because the flexibility outlined previously is also transferred to data assumptions (equality of variances, symmetry).It is also worth mentioning that some limitations occasionally cited for not using Bayesian techniques concern the lack of specialized software and high computer requirements. However, the continuous development of open-source libraries (JASP,[19] R[20]) has made it possible to estimate complex models in a Bayesian way, both in personal computers and with big data sets.

APPLICATIONS IN REHABILITATION RESEARCH

To introduce the Bayesian perspective in the data analysis of this field, a few studies have been considered. For example, a clinical trial featured in Rodríguez-Romero et al.[21] describes a therapeutic exercise program developed by a group of 19 shellfish gatherers. The level of pain, trunk muscle endurance, and fear avoidance behavior (FABQ) were tested. After sixteen 80-min sessions, the prevalence of lower and upper back pain decreased significantly. A relevant increase in the muscular endurance (in seconds) was detected in trunk extensors (TEST_EXT)(67.1 ± 42 vs. 96.1 ± 55.2, $P = 0.005$) besides a reduction in FABQ (53.9 ± 18.8) versus48.8 ± 19.7, $P = 0.09$, in a one-tailed Student's t test and 0.18 in the two-tailed test).

To compare Bayesian analysis with traditional or frequentist inference analysis, the TEST_EXT variable is used as the first example. Inference processes on results of experiments usually begin by obtaining confidence intervals about the parameters of the variables (mean, prevalence…). For the mean difference between the muscular endurance (before and after the treatment), the frequentist confidence interval is (10.1–47.8). As it is known, the confidence interval is a range of values generated by a procedure that, on repeated sampling, has a fixed probability of containing the parameter. There is no way of knowing which values have more probability, and the interval calculated from a particular sample does not necessarily include the true value of the corresponding parameter.[22]

An equivalent Bayesian concept is the CI.[16] Instead of considering the unknown parameters as fixed quantities, the Bayesian approach considers them as random variables. In this way, researchers can express their confidence for certain values for a parameter (e.g., the mean pain level, the effect size), by considering it as a random variable with some (prior) distribution. This can be a totally objective (uninformative) distribution (as in the frequentist case) or may incorporate some clinician's prior beliefs based on his/her experience.[13,23]

In our case (example 1), to obtain a Bayesian CI for the mean difference parameter, a totally objective prior distribution was considered. Thus, no preference for the value of this parameter (which would be possible if we had the results of any previous trial) was expressed (see the specific details in Rouder et al.[24]).

The posterior distribution (represented by a histogram and a smoothed polygon of frequencies) is displayed in Figure 1, together with the 95% CI of 8.7 to 47.0, where the most probable value (the maximum of the function) is near 27. Computations and Figure 1 were done with the free software R.[20]

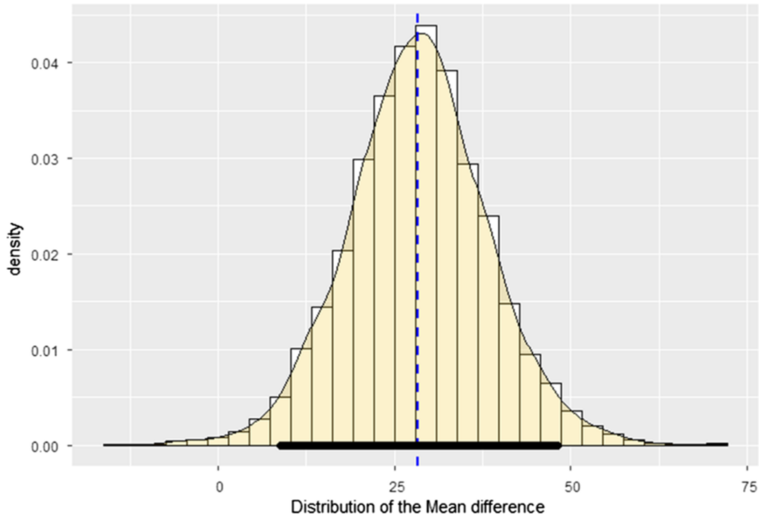

FIGURE 1. Muscular endurance test (example 1): Histogram and density estimation of the posterior distribution of the mean differences. The CI with probability 0.95 takes values from 8.7 to 47. This interval is a little bit larger than the confidence interval (10.1 to 47.8), but the distribution allows us to see which values have more probability (represented by higher values in the curve). Therefore, it is observed that extreme values have a much lower probability. These kinds of conclusions cannot be obtained with the classical confidence intervals.

Figure 1 shows the shape of the distribution of the mean difference and those values that have a higher or lower probability. Here, the advantage of the Bayesian analysis is evident because, despite not incorporating any subjective information, results are more descriptive than in the traditional method. In Table 1, the numerical information of this example and the following ones can be consulted for simplifying the comparison.

TABLE 1. Comparision between frequentist and Bayesian results in the considered examples

Example	Confidence Interval ("Frequentist")	*P*	CI (Bayesian)	BF
1	(10.1 to 47.8)	0.005	(8.7 to 47.0)	9.95
2	(−∞ to +0.07)	0.09	(−0.97 to −0.02)	0.95
3		0.044−0.018		1.998−4.122

Columns 2 and 3 correspond to confidence intervals and *P* values, and columns 4 and 5 correspond to Cis and BF

HYPOTHESIS TESTING

The following stage, in many statistical analyses, is concerned with hypothesis testing. A null hypothesis significance test consists of evaluating whether a treatment has some real effect (hypothesis H_1), rather than supporting the null hypothesis H_0 (i.e., the effect does not exist). A null hypothesis significance test has some limitations because, for example, it strongly depends on the sample size.[25] Frequentist tests guide the researcher's decision through *P* values, which only tells us if the null hypothesis can be rejected or not (according to the data observed), and it ignores what can be expected if the alternative hypothesis were true. Moreover, the P value has a number of theoretical shortcomings producing a lot of misuses and misinterpretations.[2,11]

The direct alternative tool to the P value in a Bayesian statistical test is the named BF,[26,27] computed by

Where $P(\text{Data}/H_1)$ expresses the probability of the experimental data conditioned to hypothesis H_1, (the same for H_0), that is, the probability of obtaining these data when the corresponding hypothesis is true. The BF_{10} quantifies the evidence provided by the data in favor of one model (H_1) as opposed to the other (H_0). For example, if $BF_{10} = 5$, the data are five times more likely under H_1 than under H_0.

Again, Bayesian methods can make use of any initial clinician's experience, by stating prior distributions both for the null hypothesis and the alternative, and this information can be updated by the obtained data. On the other hand, if no initial information is known, it can be proven that a Bayesian test would provide at least the same statistical information as a null hypothesis significance test.[11] Several Bayesian tests have been designed for evaluating classical comparisons (Student's t test, analysis of variance, etc.).[5] Furthermore, some of them allow less restrictive conditions about the sample data than the typical procedures (the normality assumption is not necessary, and atypical data are allowed).[12]To illustrate these concepts, we use the variable muscular endurance test again and the variable FABQ. In the first case, significant differences were found before and after the treatment, by means of the classical Student's paired t test for checking mean differences (P = 0.005). For this case, a BF_{10} of 9.95 was obtained. This value is close to strong evidence in support of H_1 (see the table of Kass and Raftery [17] with the classification of the values for the BF). Figure 2 shows a descriptive graph of the BF as a function of the sample size. This enables us to assess the robustness of this statistical factor providing valuable information for the design of future trials.

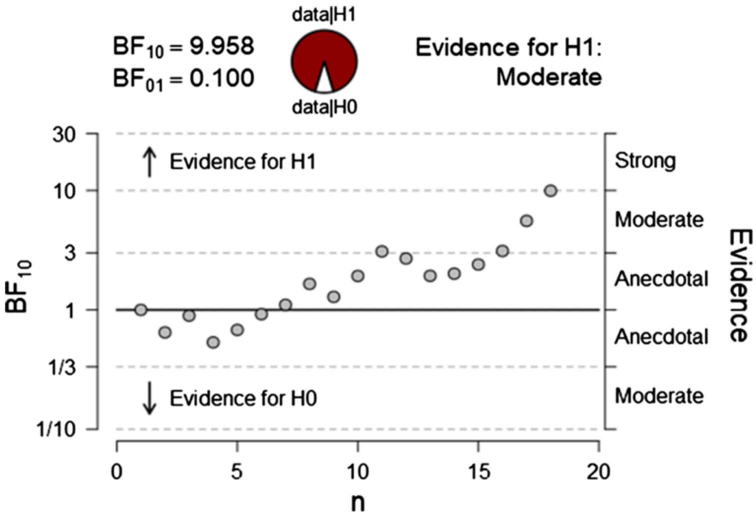

FIGURE 2. Example 1: BF as a function of the sample size. Up to a sample size of 10, evidence is approximately 1, and next the BF is increasing from anecdotal to almost strong evidence for the alternative hypothesis, indicating that with larger sample sizes, the reliability of this hypothesis will become even stronger.

The same kind of analysis was developed for the variable FABQ (Example 2). Here, the results showed a reduction of the mean value (FABQ) before and after the treatment, but, in this case, the hypothesis test for the mean differences threw a P value of 0.09 in a one-tailed Student's t test (0.18 in the two-tailed test). Therefore, the statistical significance is not clear, and thus, it does not support the practical importance of the treatment.

For a simple way of quantifying the real differences in practice, it is pertinent to perform a hypothesis test for the effect size. Formally, the null hypothesis is H_0: [delta] = 0, (null effect) and the alternative asserts a true effect ([delta] < 0), where [delta] is the mean difference divided by the standard deviation. In this case, the BF is BF_{10} = 0.95 (not producing enough evidence in favor of the alternative). However, the 95% CI is (-0.971 to -0.026), which only has negative values, because fear avoidance behavior has decreased after the intervention. Therefore, it is expected that the effect size could be greater with a larger sample size. On the contrary, the one-sided 95% confidence interval for the effect size computed in the frequentist case is (-[infinity], +0.077), which is obviously highly inaccurate and contains positive values (Table 1).

In example 3, we consider a different study in the rehabilitation field. This study examined the effects of motor imitation training using virtual reality in patients with Parkinson disease.[28] Movement features and changes in corticospinal excitability (measured by transcranial magnetic stimulation) were evaluated after the therapy in both an experimental (n = 8) and a control group (n = 7). A virtual reality-based imitation protocol was applied to a group of patients experiencing Parkinson disease. Patients performed a repetitive finger tapping task imitating the avatar during 4 wks (experimental group), whereas in a control group, the same protocol was performed without imitation. The silent period (a pause in electromyographic activity) was measured before (pre-value), immediately after finishing the 4 wks (post-value) training, and at 2 wks from the end (follow-up value). Silent period was elicited by a single transcranial magnetic stimulation pulse delivered on the motor cortex during muscle contraction.

An analysis of variance of repeated measures was performed to evaluate the effects between groups after the intervention. A significant change in the experimental group toward a more physiological profile (increased silent period and amplitude of movement) was observed. For the silent period, the observed P value was 0.045, very close to the limit level for rejecting the null hypothesis. When a Bayesian analysis of variance was performed, a BF of 1.998 was obtained, which supports a model in favor of the treatments' efficacy. Moreover, in the post hoc test, a BF of 4.122 was obtained in a test comparing the pre-value and the post-value (resulting in substantial evidence in favor of the treatment's efficacy, according to Kass and Raftery [17]). Therefore, Bayesian analysis confirms the conclusions obtained with traditional methods in Robles-Garcia et al.[28]

CONCLUSIONS

This article shows some examples in the rehabilitation field, which demonstrate the advantages of the Bayesian approach to implement different statistical analysis such as inference, hypothesis testing, and a design of repeated measures. Example 1 shows that the Bayesian CI and the graphical display of the posterior distribution bring in profits for the statistical information of results. Example 2 demonstrates that accurate conclusions of a classical null hypothesis significance test by means of a P value closer to 0 can be complemented by a high BF. Moreover, the Bayesian CI validates the existence of a practical enhancement with the treatment, whereas the classic results could be ambiguous. Finally, example 3, developed in a different context, once again supports the results in a study with a small sample size. These cases exemplify specific situations in the rehabilitation and physical therapy discipline, but representative ones. In any case, more complex situations could also be treated by means of this methodology. Studies such as meta-analyses,[29] hierarchical,[30] linear, and nonlinear models[10] can be also analyzed under a Bayesian perspective, supporting the Bayesian methods in the rehabilitation field.

ACKNOWLEDGMENTS

We thank an anonymous associate editor for sharing many useful suggestions.

REFERENCES

1. Indrayan A, Malhotra RK: Medical Biostatistics. New York: Chapman and Hall/CRC; 2017
2. Goodman S: A dirty dozen: twelve p-value misconceptions. Semin Hematol 2008;45:135–40
3. Nuzzo R: Scientific method: statistical errors. Nature 2014;506:150–2
4. Gurrin LC, Kurinczuk JJ, Burton PR: Bayesian statistics in medical research: an intuitive alternative to conventional data analysis. J Eval Clin Pract 2000;6:193–204
5. Gelman A, Rubin DB, Stern HS, et al: Bayesian Data Analysis. New York: Chapman and Hall/CRC; 1995
6. Ashby D, Smith AF: Evidence-based medicine as Bayesian decision-making. Stat Med 2000;19:3291–305
7. Burton PR, Gurrin LC, Campbell MJ: Clinical significance not statistical significance: a simple Bayesian alternative to P values. J Epidemiol Community Health 1998;52:318–23
8. Berger JO, Pericchi LR: The intrinsic Bayes factor for model selection and prediction. J Am Stat Assoc 1996;91:109–22
9. Barnett SD, Heinemann AW, Libin A, et al: Small N designs for rehabilitation research. J Rehabil Res Dev 2012;49:175
10. Broemeling LD: Bayesian Analysis of Linear Models. New York: Marcel Dekker; 1984
11. Kruschke JK: Bayesian estimation supersedes the t test. J Exp Psychol Gen 2013;142:573–603
12. Rouder JN, Haaf JM, Vandekerckhove J: Bayesian inference for psychology, part IV: parameter estimation and Bayes factors. Psychon Bull Rev 2018;25:102–13
13. Nuzzo RL: An Introduction to Bayesian Data Analysis for Correlations. PM&R 2017:9:1278–82
14. Yuen K-V: Bayesian Methods for Structural Dynamics and Civil Engineering. John Wiley & Sons Asia; 2010. Available at: https://www.wiley.com/en-us/Bayesian+Methods+for+Structural+Dynamics+and+Civil+Engineering-p-9780470824542
15. Thompson DR, Martin CR: Bayes' Theorem and Its Application to Cardiovascular Nursing. Vol 16. London, England: SAGE Publications Sage UK; 2017
16. Haskins R, Osmotherly PG, Tuyl F, et al: Uncertainty in clinical prediction rules: the value of credible intervals. J Orthop Sport Phys Ther 2014;44:85–91
17. Kass RE, Raftery AE: Bayes factors. J Am Stat 1995;90:773–95
18. Van De Schoot R, Broere JJ, Perryck KH, et al: Analyzing small data sets using Bayesian estimation: the case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. Eur J Psychotraumatol 2015;6:25216
19. Marsman M, Wagenmakers EJ: Bayesian benefits with JASP. Eur J Dev Psychol 2017;14:545–55
20. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, 2018. Vienna, Austria. Available at: https//www.R-project.org
21. Rodríguez-Romero B, Martínez-Rodríguez A, Pita-Fernández S, et al: Efficacy of a multimodal therapeutic exercise program in shellfish gatherers for the prevention of musculoskeletal disorders: a quasi-experimental study. J Sports Med Phys Fitness 2011;51:616–24
22. O'Brien SF, Yi QL: How do I interpret a confidence interval? Transfusion 2016;56:1680–3
23. Wetzels R, Raaijmakers JG, Jakab E, et al: How to quantify support for and against the null hypothesis: a flexible WinBUGS implementation of a default Bayesian t test. Psychon Bull Rev 2009;16:752–60
24. Rouder JN, Speckman PL, Sun D, et al: Bayesian t tests for accepting and rejecting the null hypothesis. Psychon Bull Rev 2009;16:225–37
25. Anderson DR, Burnham KP, Thompson WL: Null hypothesis testing: problems, prevalence, and an alternative. J Wildl Manage 2000;64:912
26. Goodman SN: Toward evidence-based medical statistics. 2: the Bayes factor. Ann Intern Med 1999;130:1005–13

27. Smith AF, Spiegelhalter DJ: Bayes factors and choice criteria for linear models. J R Stat Soc Ser B 1980;42:213–20
28. Robles-García V, Corral-Bergantiños Y, Espinosa N, et al: Effects of movement imitation training in Parkinson's disease: a virtual reality pilot study. Parkinsonism Relat Disord 2016;26:17–23
29. Sutton AJ, Abrams KR: Bayesian methods in meta-analysis and evidence synthesis. Stat Methods Med Res 2001;10:277–303
30. Langseth H, Nielsen TD: Classification using hierarchical naive bayes models. Mach Learn 2006;63:135–59