

Bayesian statistics has long been overlooked in social policy research. Over the last 25 years, there has been a renaissance in the development and application of Bayesian statistical methods, due mostly to the increased availability of powerful statistical software tools. As a result, Bayesian methods are now a viable alternative to frequentist (conventional) statistics for use in social policy research. The goal of this brief is to introduce practicing policy researchers and evaluators to the Bayesian perspective while highlighting potential advantages of Bayesian inference over frequentist inference. This brief also highlights potential next steps for advancing the use of Bayesian methods in social policy research.

Acknowledgments

I would like to thank Rebekah Selekman of Mathematica Policy Research, and Jenessa Malin and Marie Lawrence from the Administration for Children and Families' (ACF) Office of Planning, Research, and Evaluation (OPRE) for valuable comments on this brief.

David Kaplan is the Patricia Busk Professor of Quantitative Methods, Department of Educational Psychology, University of Wisconsin Madison.

KEY DIFFERENCES BETWEEN FREQUENTIST AND BAYESIAN STATISTICS

The major differences between the frequentist and Bayesian schools of statistical inference concern (1) probability and (2) unknown parameters. Both have immediate relevance for social policy analysis.

Probability. For frequentists, probability is best represented by the example of a coin toss. A frequentist analysis would examine the long-run probability of a coin landing on “tails” after repeated coin tosses. In the same way, all our conclusions regarding the results of a randomized controlled trial (RCT) presume that the same social policy experiment can be conducted a very large number of times under exactly the same conditions. In contrast, the Bayesian view of probability relies on subjective belief and is best represented by a “bet.” In social policy research, the bet might be stated as, *“Based on past research, what is the probability that my proposed intervention will have an effect greater than 15 points over a control?”* In this example, probability is based not on an infinite number of repeatable and independent experiments but rather on how much prior knowledge the researcher has

about past similar interventions and how precise the researcher believes their prior knowledge to be.

Unknown parameters. Another key difference between the frequentist and Bayesian schools concerns the nature of the unknown parameters, such as the true impact of a proposed intervention in an RCT. Frequentists assume that the true treatment effect is unknown and that the estimation of the treatment effect does not incorporate our beliefs about reasonable values that the effect can take on. In Bayesian statistics, the treatment effect is also considered unknown, but this treatment effect must now be described by a probability distribution shaped by the researcher's best guess about the size of the treatment effect and their uncertainty about that guess. This probability distribution describing the researcher's best guess is referred to as the prior distribution, or simply, the prior.¹

INCORPORATING PRIOR KNOWLEDGE TO ADDRESS POLICY-RELEVANT QUESTIONS

Bayes' theorem² is a model for learning from data. For example, consider an experiment designed to examine whether a particular intervention embedded in a human services program improves child and family outcomes. Like frequentists, Bayesians begin by

¹ Note that for more complex models involving many variables, each parameter is required to have a prior distribution.

² For a basic review of Bayes' theorem, see Kaplan, D. (2014). *Bayesian statistics for the social sciences*. New York, NY: Guilford Press.

assuming that the effect of the intervention can be described by a probability distribution defined by a set of parameters. For example, we may assume that the outcome of interest is normally distributed and modeled as a function of the intervention (perhaps in a simple dummy variable regression where 1 = the intervention group and 0 = business-as-usual group). This model for the outcome is referred to as the data distribution. The outcome of interest is the difference among those individuals assigned to the intervention and those individuals assigned to business as usual.

Unlike in frequentist analysis, Bayesian analysis calls on researchers to directly incorporate any prior knowledge they have about the intervention effect by using the prior distribution. Using Bayes' theorem, researchers weight their prior beliefs by the observed data through multiplying the prior distribution by the data distribution. The posterior distribution that results from the multiplication of the prior distribution by the data distribution is an updated estimate of the intervention effect. From the posterior distribution, a researcher can easily calculate a point estimate of the treatment effect and a "credible interval" for the point estimate. With the point estimate and the credible interval in hand, the research can address such questions as, "*What is the probability that the effect of the intervention is greater than zero?*" or, "*What is the probability that the effect of the intervention is between the values of X and Y?*" where X and Y might be values of policy relevance. These types of questions cannot be answered from the frequentist perspective.

How to identify and use priors in

practice. A prior distribution is nothing more than a probability distribution itself and is characterized by its own set of parameters. The values of these parameters are set by the researcher or evaluator reflecting their belief about the true value of the treatment effect and their uncertainty around that value. In principle, these estimates can come from any source. In social policy research, prior information might come from relevant past studies, meta-analytic studies, and/or expert opinion. The use of prior distributions is what separates Bayesian analysis from frequentist analysis. Researchers use prior distributions to express their uncertainty about the effect of an intervention in rigorous statistical terms.

In general, there are three types of priors:

1. ***Noninformative priors.*** Researchers use noninformative priors to reflect actual or feigned ignorance about the treatment effect. Actual ignorance might be present in cases where the intervention is being conducted for the very first time, and nothing is known about the values that the intervention effect could take on. Feigned ignorance might occur when researchers want to refrain from asserting an opinion about the average value and precision of the intervention effect. A noninformative prior distribution is sometimes referred to as “flat,” reflecting the researcher’s view that the treatment effect could take on any value.
2. ***Weakly informative priors.*** Researchers use weakly informative priors to prevent parameters from taking on values that are known to be mathematically or physically impossible. For example, a weakly informative prior might be placed

on a variance term so that it cannot be negative—indicating that the intervention effect must be positive or zero—but no more information is provided about the intervention effect.

3. ***Informative priors.*** Researchers use informative priors when they have access to cumulative evidence or expert opinion about the treatment effect that can reasonably apply to the intervention under study. For example, previous information from prior meta-analytic studies might indicate the average effect of an intervention across multiple studies. The precision placed around the average conveys the researcher’s degree of certainty around that effect. The lower the precision, the greater the expression of uncertainty.

ADVANTAGES OF BAYESIAN ANALYSIS

Null hypothesis significance testing is the dominant method of statistical inference in social policy research. However, it has sustained years of well-documented criticism. Research consumers (and many researchers themselves) often misinterpret the p -value from significance tests, confusing it with Type 1 error, or the probability of claiming an effect when there is none. In addition, significance tests often lack policy-relevant interpretation. Significance tests represent the likelihood of seeing results at least as extreme as the observed effect. When evaluating an intervention, the p -value represents the likelihood of seeing the observed effect or any effect greater than what is observed, assuming the true effect is zero. This finding may be difficult to interpret

in a policy context. In contrast, a Bayesian analysis would indicate the likelihood that an intervention produces an effect greater than a specific number, such as zero or within a specified range of policy-relevant values. Using Bayesian methods, a researcher can directly answer the question, “*What is the probability that the intervention had a positive effect?*”

A researcher can also form a posterior probability interval. For example, when evaluating an intervention, a 95-percent posterior probability interval would mean the probability that the intervention effect lies in the interval is 95 percent. In contrast, a frequentist confidence interval represents the range of values between which the observed intervention effect would lie in 95 percent of repeated studies. Again, the Bayesian posterior probability interval may be easier to understand and more informative for addressing real-world policy questions.

ADVANCING THE USE OF BAYESIAN METHODS IN SOCIAL POLICY RESEARCH

Despite the advantages and recent accessibility of Bayesian methods, they are not commonly used in social policy research. The following three steps would help advance Bayesian methods in social policy research:

1. Extend methodological requirements in grant and contract proposal submission guidelines.
2. Develop requirements for evaluating publications that use Bayesian methods.

3. Adapt requirements for systematic evidence review submissions.

Grant and contract proposal submission guidelines. Federal agencies and foundations that award research contracts and grants often require that proposals provide a detailed discussion of the statistical methods to be employed in each proposed project. For example, a competitive proposal typically requires a discussion of sample size, effect sizes, and power analyses. These requirements are situated within the frequentist methods of statistics, and in particular, null hypothesis significance testing. Except where prohibited by law, proposal submission guidelines related to statistical methodology should be updated to include and define high-quality implementation of Bayesian methods.

Requirements for publications. Given the increasing use of Bayesian methods in briefs, reports, and other publications resulting from Federal grants or contracts, clear guidelines for the communication of these methods are needed. For example, Federal project officers could ensure that findings contain appropriate information about key assumptions and methodological choices while clearly communicating findings to the intended audience.

Systematic evidence reviews. An increasing number of Federal agencies support systematic evidence reviews, including the U.S. Department of Education’s (ED) What Works Clearinghouse³ (WWC) and the U.S. Department of Health and Human Services’ (HHS) Home Visiting Evidence of Effectiveness⁴ and Employment

³ See <https://ies.ed.gov/ncee/wwc/>.

⁴ See <https://homvee.acf.hhs.gov/>.

Strategies for Low-Income Adults Evidence Review.⁵ These systematic evidence reviews are typically established within the frequentist framework of significance testing, often assessing study designs and describing study effects using cutoffs based on statistical significance and associated effect sizes. Systematic evidence reviews should establish guidelines for high-quality Bayesian evidence. For example, Mark Schneider, director of ED's Institute of Education Sciences (IES), has indicated that the WWC will soon update practices with regard to statistical significance.⁶ He stated, *"The use of $p = <.05$ as a cutoff for statistical significance is a convention of long standing. Yet modern thinking and many new statistical approaches have outdated that bright line test. We will be working on how to reflect these changes in how we judge research in the WWC. As we do that, we will be ending the 'substantively important' designation."*

GUIDELINES FOR JUDGING THE USE OF BAYESIAN METHODS

These steps require rigorous guidelines for judging the quality of work that seeks to use Bayesian methods. The following four guidelines are necessary (but perhaps not sufficient) requirements for competitive proposals, publications, and/or submissions to systematic evidence reviews in which Bayesian methods are used.

⁵ See <https://employmentstrategies.acf.hhs.gov/>.

⁶ U.S. Department of Education, Institute of Education Sciences. (n.d.). *Changes are coming to What Works Clearinghouse* [Statement by IES Director Mark Schneider]. Retrieved from <https://ies.ed.gov/director/wwc2018.asp>.

Model specification. Using Bayesian statistical methods requires that information be provided about the probability distribution of the outcome of interest and prior probability distributions for all model parameters. Researchers should provide these distributions along with the usual requirements of the design specifications.

Source of prior information. Closely related to the need for clear model specification is a detailed discussion of the process by which priors were constructed. For example, if researchers obtained priors through expert group panels, they should provide a detailed discussion of how the panels were organized and the methods for eliciting priors from experts. If researchers obtained priors from meta-analytic or prior research studies, they should identify these studies and make them accessible to the larger research community.

Computational considerations. Bayesian analysis is computationally intensive, and the results obtained from a Bayesian analysis are not interpretable unless there are assurances that the computational algorithms have worked properly. Fortunately, there are many diagnostic measures available in Bayesian software that allow researchers to present evidence that the computations have properly stabilized. Researchers should provide these diagnostic results as evidence of computational stability so reviewers can evaluate the interpretability of the results.

Sensitivity analyses. A common concern regarding the use of Bayesian statistical methods is the sensitivity of the results to the researcher's choice of prior distributions, especially in cases in which the sample size

is relatively small. In those cases, minor changes to the priors can impact the results. In cases in which the sample size is large, the results of a Bayesian analysis will likely be quite similar to those of a frequentist analysis, although the interpretations will be different. Regardless, it is essential that researchers propose or conduct some form of a sensitivity analysis comparing the specified priors with other reasonable priors and even noninformative priors. A robust analysis will demonstrate that the choice of priors produces very little change in the results.

CONCLUSION

Bayesian statistical inference represents a useful and increasingly common tool for creating and updating knowledge relevant to social policy research. Bayesian inference also offers an answer to the challenges of interpretability and policy relevance inherent in null hypothesis significance testing. Government leaders have a critical role to play in encouraging the adoption of Bayesian methods in grant and contract proposals, publications, and systematic evidence reviews.

This brief was prepared for the Office of Planning, Research, and Evaluation (OPRE), part of the U.S. Department of Health and Human Services' (HHS) Administration for Children and Families. It was developed under Contract Number HHSP233201500109I. Insight Policy Research (Insight), located at 1901 North Moore Street, Suite 1100 in Arlington, Virginia, assisted with preparing and formatting the brief. The ACF project officer is Emily Ball Jabbour, and the ACF project specialist is Kriti Jain. The Insight project director is Rachel Holzwart, and the Insight deputy project director is Hilary Wagner.

This brief is in the public domain. Permission to reproduce is not necessary. Suggested citation: Kaplan, D. (2019). *Bayesian inference for social policy research* (OPRE Report 2019 36). Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

This brief and other reports sponsored by OPRE are available at www.acf.hhs.gov/opre.

Disclaimer: The views expressed in this publication do not necessarily reflect the views or policies of OPRE, ACF, or HHS.



[Sign-up for the ACF
OPRE News E-Newsletter](#)



Like OPRE on Facebook
Facebook.com/OPRE.ACF



Follow OPRE on Twitter
[@OPRE_ACF](https://twitter.com/OPRE_ACF)