Bayesian Methods for Social Policy Research and Evaluation

OPRE Report #2018-38
Holzwart and Wright, Insight Policy Research
March 2018

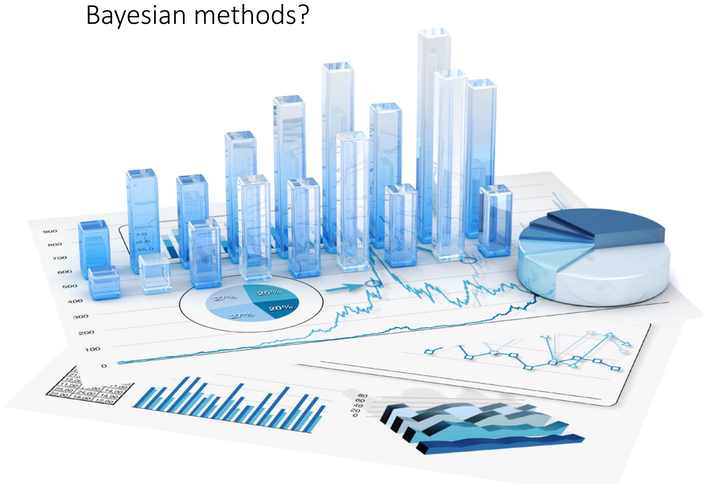
INTRODUCTION

Probability (*p*) values are widely used in social science research and evaluation to guide decisions on program and policy changes. However, as summarized in a 2016 statement from the [American Statistical Association (ASA)][1], *p*-values have some inherent limitations that may lead to misuse, misinterpretation, or misinformed decisions.

Bayesian methods, which use probabilistic inference to determine the importance of a finding, are becoming the primary alternative approach to *p*-values. Given the increasing attention to and use of Bayesian methods in social science research, it is essential to understand the underlying assumptions, tradeoffs, validity, and generalizability of results in a Bayesian framework, and the circumstances under which there may be advantages to using a Bayesian rather than, or in addition to, a frequentist approach.

On October 19 and 20, 2017, OPRE brought together a diverse group of participants from Federal agencies, research firms, foundations, and academia to discuss Bayesian methods for use in social policy research and evaluation. The meeting addressed several key questions:

1. What are the advantages and disadvantages of Bayesian methods?
2. How are Bayesian models built and evaluated?
3. How have Bayesian methods been used successfully in the field to date?
4. How can social policy researchers and stakeholders use Bayesian methods to better support decision-making?



This brief summarizes key themes that emerged from the meeting.

> **The Bayesian approach is best represented by a "bet." Probability is not based on an infinite number of events (as it is in a frequentist approach), but rather on how much knowledge someone has about a potential event and how much they are willing to stake on that knowledge.**
>
> —Summarized from David Kaplan's "A Brief Introduction to Bayesian Statistics" presentation

WHAT ARE THE ADVANTAGES AND DISADVANTAGES OF BAYESIAN METHODS?

Presenters at the meeting discussed the advantages and disadvantages of using Bayesian methods in a social policy context. One advantage relative to frequentist methods is that Bayesian results, which are presented in probabilistic terms, can be more useful to policymakers. Frequentist approaches use strict cutoff values (e.g., $p<0.05$) to answer the question "If the unobserved true effect is zero, how likely is it to obtain a result as extreme as the one observed?" This is rarely the question a policymaker actually wants to answer. Bayesian methods, on the other hand, answer the question "What is the probability that the impact of X is at least Y," where Y is the impact meaningful to the policymaker?

[1] American Statistical Association (ASA) (2016), "ASA Statement on Statistical Significance and P values," The American Statistician, 70, DOI: 10.1080/00031305.2016.1154108.

This approach can be more intuitive, particularly to lay audiences without formal statistical training. Other strengths of Bayesian methods include the ability to:

- Quantify multiple types of uncertainty (frequentist approaches such as *p*-values, only account for sampling uncertainty)
- Incorporate existing information (either from prior studies or from the dataset at hand) and update conclusions as new data become available
- Improve precision in studies with small samples (or small subgroups) and low power
- Better evaluate models, as Bayesian model-building requires researchers to specify their decisions and is therefore more transparent than frequentist methods

There are also challenges to using Bayesian methods in a public policy context. For example, it can be difficult to translate Bayesian probabilistic inferences into the yes/no responses often required for policy decisions. Bayesian models are also computationally intensive. Until recent advances in computing (such as increases in processor power and advances in distributed processing environments), the time required to run Bayesian models was impractical for most applications. Building Bayesian models requires researchers to select prior assumptions, but selecting an appropriate set of prior assumptions can be challenging. Further, many researchers lack the knowledge and training to confidently implement a Bayesian analysis, even when a p-value may not be the most useful approach for a given question or dataset, or when a Bayesian approach is more efficient and effective.

HOW ARE BAYESIAN MODELS BUILT AND EVALUATED?

A key advantage of Bayesian methods is that they allow researchers to combine estimates from study data with relevant outside information (a prior probability distribution), to derive a posterior distribution. The prior probability distribution can reflect all information available to date on a model parameter.

When developing and evaluating Bayesian models, researchers should:

- Carefully select priors, drawing on a population of similar studies to the one being conducted and reviewing the distribution of impacts across the studies in that population
- Be mindful that subjective priors in public policy should be well-documented and based on factual prior knowledge
- Conduct sensitivity analyses to test the extent to which the priors drive the results
- Evaluate how well a chosen model fits the data using cross-validation (assessing how results will generalize to an independent data set) and model comparison techniques that assess how closely data generated from the model matches the actual data (posterior predictive checks) or how different regression models compare (using Bayes factors)
- Focus on producing replicable results, drawing on the connection between theory, measurement, and data

HOW HAVE BAYESIAN METHODS BEEN USED SUCCESSFULLY IN THE FIELD TO DATE?

Presenters shared ways researchers have applied Bayesian methods in a public policy context, describing several examples over the course of the 2-day conference:

- Researchers at RTI used Bayesian methods to create an interactive dashboard for the Centers for Medicare & Medicaid Services (CMS)'s meta-evaluation of the Health Care Innovation Awards. The dashboard enabled decision-makers to conduct nuanced analyses by adjusting the level of risk tolerance and filtering data to answer questions

such as, "What is the probability that a particular program reduces hospital readmissions by 5 days or more?"

- Researchers at RAND applied a Bayesian approach to create statistical benchmarks for CMS's Health Care Provider Performance Evaluation. Using prior information to calculate provider performance estimates, including a weighted average of the hospital's own ('direct') estimate and the average for all hospitals, they were able to assess performance more accurately for hospitals that served a small number of patients.
- Researchers at Mathematica Policy Research used a Bayesian meta-regression to analyze findings from the Employment Strategies for Low Income Adults Evidence Review. Researchers "borrowed strength" from previous related studies and examined variation in the impacts of employment program without sacrificing precision. This permitted them to describe conclusions probabilistically using plain, intuitive language and focus on practically meaningful thresholds.
- Researchers at the University of Wisconsin used data from the Program for International Student Assessment to conduct Bayesian Model Averaging for Structural Equation Modeling (BMA-SEM). This technique takes multiple models into account rather than requiring the researcher to select a single model. BMA-SEM resulted in a model of reading proficiency that was better aligned with real world outcomes than a comparable frequentist SEM model.

HOW CAN SOCIAL POLICY RESEARCHERS AND STAKEHOLDERS USE BAYESIAN METHODS TO BETTER SUPPORT DECISION-MAKING?

Two roundtable discussions focused on communicating Bayesian findings in a policy context, challenges in using Bayesian methods, and next steps for facilitating more widespread use of Bayesian methods. In exploring how to communicate Bayesian results to policymakers—who are more accustomed to a frequentist approach—panelists touched on policy community concerns related to using subjective prior assumptions. A benefit of Bayesian approaches is that they allow prior knowledge to be incorporated into the analyses. However, the researcher must decide, often in collaboration with stakeholders, what prior information is relevant to incorporate into the analysis. Research clearinghouses, such as the What Works Clearinghouse, have the potential to provide researchers a starting point for a set of priors that is based on a wealth of studies already curated and formally reviewed.

Similarly, discussions of next steps in potentially adopting Bayesian methods in more social research focused on finding effective ways to communicate findings to decision-makers at all levels. Because Bayesian methods present results in a probabilistic framework, policymakers can incorporate their own risk tolerance in making decisions. For example, where a frequentist evaluation may conclude a program has "no discernable effect," a Bayesian evaluation might conclude "there is an 87 percent chance that the impact is at least $16." Some policymakers will embrace this framework, because they can decide whether that impact is "good enough" given the program's potential costs and benefits. However, other policymakers would prefer bright-line direction from researchers (i.e., the program works or it doesn't). Panelists suggested that Federal agencies should identify transparent, objective criteria to determine whether a program or intervention is "proven" in a Bayesian framework so that Bayesian results can be rated and incorporated into evidence reviews.

WANT TO LEARN MORE?

To access the online meeting archive, including a detailed schedule, meeting materials, and presentation slides, please visit the OPRE Innovative Methods Meeting website at www.opremethodsmeeting.org. The site also includes materials from other innovative methods meetings OPRE has organized, and it will be updated to include future meetings.

Decision makers need to understand the value Bayesian methods add, and easy ways to communicate this information to all levels of government representatives. Some policy makers want a dichotomous yes or no answer, but others are more interested in the probability that program changes will improve outcomes.

—From *Roundtable Discussion: Where Do We Go From Here?*

MEETING AGENDA

Bayesian Methods for Social Policy Research and Evaluation

October 19–20, 2017, Holiday Inn Capitol, Washington, DC

Day 1, Thursday, October 19

Welcome and Opening Remarks
Naomi Goldstein (Deputy Assistant Secretary for the Office of Planning, Research, and Evaluation)

Basics of Frequentist and Bayesian Approaches
Doctor, It Hurts When I p
Ronald Wasserstein (American Statistical Association)

A Brief Introduction to Bayesian Statistics
David Kaplan (University of Wisconsin–Madison)

Specifying the Bayesian Model and Evaluating Model Fit
Plausible Priors Precede Persuasive Posteriors
Mariel Finucane (Mathematica Policy Research)

Bayesian Model Specification (Or at Least Some of What Can Be Said About This Topic in 25 Minutes)
David Draper (University of California–Santa Cruz)

The Failure of Null Hypothesis Significance Testing When Studying Incremental Changes—What To Do About it?
Andrew Gelman (Columbia University)

Bayesian Applications I
Making Bayesian Analyses Accessible Through Visualization: Case Study, Meta-Evaluation of the Health Care Innovation Awards
Nikki Freeman (RTI International)

Applications: Health Care Provider Performance Assessment
Susan Paddock (RAND Corporation)

Roundtable Discussion: Communicating Bayesian Findings in a Policy Context
Panelists:
Stuart Buck (Laura and John Arnold Foundation), Moderator
Donald Berry (MD Anderson Cancer Center, University of Texas)
Gregory Campbell (GCStat Consulting LLC)
Timothy Day (Center for Medicare and Medicaid Innovation)
Jacob Alex Klerman (Abt Associates)

Day 2, Friday, October 20

Bayesian Applications II
The Right Tool for the Job: A Bayesian Meta-Regression of Employment and Training Studies
Lauren Vollmer (Mathematica Policy Research)

On the Utility of Bayesian Model Averaging for Optimizing Prediction: Two Case Studies
David Kaplan (University of Wisconsin–Madison)

Bayesian Inference for Sample Surveys
Trivellore Raghunathan (University of Michigan)

Roundtable Discussion: Where Do We Go From Here?
Panelists:
Robin Ghertner (Office of The Assistant Secretary for Planning and Evaluation), Moderator
Scott Cody (Project Evident)
Molly Irwin (Department of Labor, Chief Evaluation Office)
Renee Mentnech (Center for Medicare and Medicaid Innovation)
David Rindskopf (CUNY Graduate Center)

This brief was prepared by Insight Policy Research (1901 N. Moore Street, Suite 1100, Arlington, VA 22209) under Contract Number HHSP233201500109I. The ACF project officers are Anna Solmeyer and Laura Nerenberg. The Insight project director is Rachel Holzwart, and the deputy project director is Hilary Sama.

 Suggested citation: Holzwart, R., & Wright, D. (2018). *Bayesian methods for social policy research and evaluation* (OPRE Report 2018-38). 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 the Office of Planning, Research, and Evaluation are available at www.acf.hhs.gov/opre.

insight POLICY RESEARCH