Why Bayesian Statistics Might Be the Right Approach for Making Policy

APRIL 2026

Charles Michalopoulos
Pei Zhu

Conducting strong research studies is a prerequisite for evidence-based programs and policy, but interpreting study findings can be challenging. For example, views differ in how to interpret a result that is not "statistically significant." Does that mean an intervention had no effect at all? Likewise, in a study where only some findings are statistically significant, often only the ones that are statistically significant are highlighted while others are ignored or downplayed. Is that the right approach?

Despite these types of ambiguities, policymakers, funders, and practitioners must make decisions on expanding an intervention with promising effects, on abandoning or strengthening an intervention with weak effects, or on doing more research to arrive at a clearer set of findings. This document compares two statistical approaches to interpreting impact findings that can help in making decisions like these: a conventional approach that uses statistical significance and that has been the standard approach in social program evaluation for decades, and a Bayesian approach that offers a more intuitive way of interpreting impact estimates.

Both approaches require expertise in producing and interpreting findings. **However, the Bayesian approach can provide direct answers to the questions that programs, funders, and policymakers often have: Does this program work? How confident can one be that effects exceed meaningful thresholds?** The questions that the conventional approach can answer are generally less useful for making policy decisions, which suggests that

mdrc
BUILDING KNOWLEDGE TO IMPROVE SOCIAL POLICY

evaluators can help decision-makers extract more value from rigorous research by providing Bayesian interpretations of impact estimates.

The Bayesian approach can provide direct answers to the questions that programs, funders, and policymakers often have: Does this program work? How confident can one be that effects exceed meaningful thresholds?

Although there are advantages to Bayesian methods, they have not been used often in social science research, and it may not be clear to program leaders, funders, and policymakers how to interpret results from a Bayesian analysis. This document is intended to fill that gap and to provide information on how the conventional and Bayesian approaches can answer a seemingly simple question: Is an intervention effective?

Two Statistical Approaches for Interpreting an Estimated Impact

Analysts can choose either the conventional approach or the Bayesian approach to interpret an intervention's estimated impact.

- **THE CONVENTIONAL (OR FREQUENTIST) APPROACH** answers the question, "How likely is it that an intervention with no true effect would produce an estimated impact this large (or larger) by chance?" It starts by assuming the intervention has no effect (the "null hypothesis") and asks whether the data provide strong enough evidence to reject that assumption. This method became increasingly used in the first half of the twentieth century as a way of assessing whether manufacturing production runs met quality criteria.[1] It is commonly used today because most researchers learn it in graduate school, it requires relatively little computational power, and its conclusions are based solely on the data.

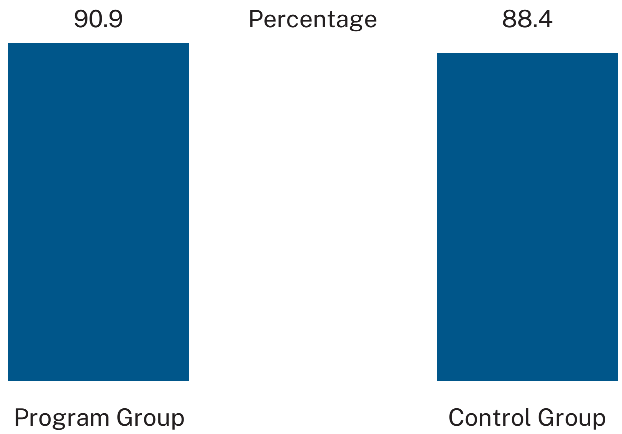
- **THE BAYESIAN APPROACH** answers questions such as, "Given the data and expectations about the likely effects of the intervention, what is the probability that it had a positive effect?" As suggested by this question, the Bayesian approach formally incorporates expectations about the likely effect of an intervention (ideally from earlier studies or, less often, from expert judgment). These expectations are referred to as a Bayesian prior. Using this approach, analysts can calculate the probability that the effect is positive or that it exceeds a specific threshold relevant to policy decisions.

An Example from an Impact Study

In MDRC's randomized trial of the Labor Force Attachment program in Grand Rapids, Michigan — which provided recipients of cash assistance with short-term job search assistance to encourage them to find employment quickly — a central question was whether the approach increased employment for participants. Results are shown in Figure 1.

Using data from state unemployment insurance records, the study found that 90.9 percent of the program group worked in the five years after they entered the study, compared with 88.4 percent of the control group. The estimated impact of 2.5 percentage points had a p-value of 0.18, which is not statistically significant.[2] (See Box 1 for the definition of a p-value and other terms.)

Figure 1. Effect of the Grand Rapids Labor Force Attachment Program on Employment



Box 1. Some Important Terms

Bayesian prior distribution: Starting assumptions about likely effects, ideally based on earlier studies or expert judgment, collected before seeing new data.

p-value: The probability of observing effects at least as large as those estimated if the true effect were zero. (Note: The p-value is not the probability that the program had no effect.)

Statistically significant: A designation applied when a p-value falls below a predetermined threshold (typically 0.05 or 0.10).

Table 1 contrasts what each approach can or cannot say about the Grand Rapids finding.

As indicated in the first two rows of the table, the p-value of 0.18 in the conventional approach *does not* mean there is an 18 percent probability that the effect is positive. That is a common misconception about significance testing. Instead, it means that there is an 18 percent probability that the study of the Grand Rapids program would produce an estimated impact of 2.5 percentage points or larger if the program were completely ineffective. The conventional approach also does not provide direct information on whether the program's estimated impact is large enough to make it the right policy to adopt.

The Bayesian approach provides a different interpretation of the evidence that may be more useful for making policy decisions. It does not address the concept of statistical significance directly but presents information about the how large the effects are likely to be. In this case, a Bayesian approach indicates there is a 91 percent to 98 percent probability that the effect of the Grand Rapids program on employment is positive and a 50 percent to 60 percent probability that it is at least as large as 2.5 percentage points. A similar probability could be calculated for any policy-relevant threshold.

The questions that the Bayesian approach can answer are likely to be of greater interest to policymakers, funders, and program leaders because they provide direct statements about the probability of the intervention's effectiveness.

Table 1: What Each Approach Can Tell Us About the Grand Rapids Labor Force Attachment Finding

QUESTION	CONVENTIONAL APPROACH	BAYESIAN APPROACH
If the program had no true effect, what is the probability that one would observe an impact estimate at least this large (2.5 percentage points)?	18 percent (the p-value multiplied by 100)	Not the focus of this approach
Is the finding "statistically significant?"	No	Not the focus of this approach
What is the probability that the program had a positive effect?	Cannot answer this question	91 percent to 99 percent
What is the probability that the effect was at least 2.5 percentage points?	Cannot answer this question	50 percent to 65 percent
What is the probability that the effect exceeded some other policy-relevant threshold?	Cannot answer this question	Can be calculated for any threshold of interest

Why Isn't the Bayesian Approach Used More Often in Impact Research?

The main barrier is not a fundamental disagreement about the value of Bayesian methods but rather training and norms. Most evaluators learn conventional hypothesis testing in graduate school, and statistical significance has been the standard in published research and funder requirements for decades. As a result, the social science research community has developed strong conventions related to p-values and significance thresholds, and its members may be less comfortable interpreting Bayesian results. Changing established practices takes time, even when newer approaches offer advantages. That said, there is growing momentum for change. Statisticians and applied researchers have increasingly voiced concerns about the limitations of significance testing, particularly its all-or-nothing threshold, its tendency to be conflated with substantive significance, and its frequent misinterpretation as the probability that an effect is real. In 2016, the American Statistical Association published a statement of concern about these issues, and a 2019 editorial in *The American Statistician* called for moving to "a world beyond $p < 0.05$." Bayesian methods offer one path forward.[3]

Historically, Bayesian analysis in impact research was also constrained by practical challenges: limited rigorous studies from which to estimate prior distributions based on strong evidence, difficulty identifying and gaining access to existing research, substantial computational requirements, and a lack of accessible software and training materials. Fortunately, these barriers have diminished significantly in recent decades. The number and quality of rigorous impact studies have grown rapidly, supported by advances in data collection, experimental and quasi-experimental research designs, and organizational capabilities at research institutions. Clearinghouses such as the What Works Clearinghouse and the Pathways to Work Clearinghouse now make it easier to identify and gain access to existing studies. Meanwhile, substantial increases in computing power, new software tools, and emerging methodological research on combining findings across studies have made Bayesian analysis increasingly feasible for applied researchers.

What Are the Advantages of the Bayesian Approach?

The Bayesian approach has some important advantages to consider:

- It is easier to interpret findings. Saying there is a 91 percent probability that the Grand Rapids Labor Force Attachment program increased employment is intuitive and clear. Saying the p-value of the estimated impact is 0.18 is not as intuitive or clear, and is frequently misinterpreted.

- It can incorporate information from other sources, such as other studies, theory, or expert opinion. Studies are not done in a vacuum. Although literature reviews have always provided context in interpreting a finding, the Bayesian approach provides a formal way of incorporating that information.

- It is not subject to the knife's edge of statistical significance. When significance testing is used, a result that is just barely statistically significant (for example, $p = 0.04$) is often treated very differently from one that is just barely not statistically significant (for example, $p = 0.06$), even if the two estimates are close. In contrast, a Bayesian approach is likely to result in similar conclusions from the two estimates. The Grand Rapids study can be used to illustrate this idea. As discussed earlier, results using unemployment insurance records found that the program increased employment by 2.5 percentage points, but that the estimate is not statistically significant. According to survey data, however, the Grand Rapids program increased employment by 2.8 percentage points, and that result is statistically significant. The conclusion from a standard analysis might depend on which data source is used, since one result is statistically significant and one is not. However, results from a Bayesian analysis using unemployment insurance records indicate a 91 percent to 98 percent probability that the Grand Rapids program increased employment, and results using surveys indicate a 97 percent to 98 percent probability that it increased employment. There is no ambiguity in these results: Both suggest substantial evidence that the Grand Rapids program increased employment.

More Information About the Bayesian Approach

This section discusses how Bayesian analyses interpret impact findings and the advantages and limitations of Bayesian analyses.

Why Is the Bayesian Result Reported as a Range Rather than One Number?

The Bayesian approach can incorporate information (called a “Bayesian prior distribution”) about the likely effects of the intervention being studied. Different Bayesian prior distributions will lead to different findings once a new study’s results are included. A prior distribution can be neutral, assuming nothing at all. For example, for the Grand Rapids study, a prior distribution that does not incorporate information from other studies — so that the Bayesian results stem solely from new data collected — produces a result indicating a 91 percent probability that the Grand Rapids program increased employment. However, the larger study that included Grand Rapids had two other Labor Force Attachment programs, in Atlanta and Riverside. On average, those two programs had slightly larger effects on employment than the one in Grand Rapids did. Using a Bayesian prior distribution that incorporates information from the two other programs produces a result indicating a 98 percent probability that the Grand Rapids program increased employment. Using multiple prior distributions is a common approach to assess whether the results are sensitive to the assumptions being made.

Where Does a Bayesian Prior Distribution Come from?

- When similar studies have been conducted in the past, a Bayesian prior distribution could be based on results from those studies. Although other studies are a typical source of information for a Bayesian prior distribution, caution is in order if the interventions or studies differ. Choosing which studies to include in developing the Bayesian prior distribution is as much art as science.

- When few earlier studies exist, a Bayesian prior distribution could be formed by asking a panel of experts about the likely effect of an intervention. This practice has been used in research on medical interventions.

- When there is little or no information on which to base expectations about a program’s effectiveness — not even an expert panel — researchers might use a “weakly informative” prior distribution. Such a prior distribution might place an equal weight on a wide range of outcomes. When a weakly informative prior distribution is used, conclusions depend primarily on the new data that are collected.

How Can a Bayesian Analyst Guard Against Choosing a Specific Prior Distribution to Get a Desired Result?

- A common concern is that incorporating the Bayesian prior distribution adds subjectivity to the analysis. An analyst can generate a positive finding by starting with an unrealistically positive Bayesian prior distribution, or can generate a negative finding by doing the opposite. This concern is especially relevant when combining the Bayesian prior with data from a new, small-scale study.

- One way to guard against this possibility is to base Bayesian prior distributions on external evidence, such as results from other studies or a meta-analysis of similar programs. Another way is to test the sensitivity of results to reasonable alternative Bayesian prior distributions.

- It can also help to be transparent about where the Bayesian prior distribution comes from, to discuss what it implies before looking at the study results, and to report it alongside the main findings. If possible, the source of the Bayesian prior distribution could be included in a study's preregistered analysis plan.

- Such subjectivity is not unique to the Bayesian approach; two people might draw different conclusions from the same finding using conventional methods if they bring their own evidence to interpreting the finding.

The Bottom Line

Both conventional and Bayesian approaches require expertise to produce and interpret findings. However, the Bayesian approach can increase the usefulness of evaluation findings by providing direct answers to the questions programs, funders, and policymakers most often have: What is the probability that this program worked? How confident can one be that effects exceed meaningful thresholds? Evaluators can help decision-makers extract more value from rigorous research by providing Bayesian interpretations of impact estimates. Funders can move the field forward by prompting researchers to provide these interpretations.

Notes and References

1. Julien Gradoz, "Promoting Statistical Methods to Engineers and Economists: Walter Andrew Shewhart and Statistical Quality Control," CHOPE Working Paper No. 2025-02 (Duke University, Center for the History of Political Economy [CHOPE], 2025).
2. Gayle Hamilton, Stephen Freedman, Lisa A. Gennetian, Charles Michalopoulos, Johanna Walter, Diana Adams-Ciardullo, Anna Gassman-Pines, Sharon McGroder, Martha Zaslow, Jennifer Brooks, and Surjeet Ahluwalia, *How Effective Are Different Welfare-to-Work Approaches? Five-Year Adult and Child Impacts for Eleven Programs* (U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation and Administration for Children and Families, and U.S. Department of Education, 2001).
3. Ronald L. Wasserstein and Nicole A. Lazar, "The ASA Statement on P-Values: Context, Process, and Purpose," *American Statistician* 70, 2 (2016), 129–133, http://doi.org/10.1080/00031305.2016.1154108; Ronald L. Wasserstein, Allen L. Schirm, and Nicole A. Lazar, "Moving to a World Beyond 'p<0.05'," *American Statistician* 73, Sup. 1 (2019): 1–19, https://doi.org/10.1080/00031305.2019.1583913.

Acknowledgments

The authors are grateful to Carolyn Hill for her encouragement and her collaboration in shaping the document from inception to publication. We also would like to thank Barbara Condliffe, Shira Mattera, and Leigh Parise at MDRC for their thoughtful comments on an early draft. Joshua Malbin edited the brief and Carolyn Thomas prepared it for publication. Finally, we would like to thank Dan Tsin and Blair Hewes at Blue Meridian Partners for helpful comments.

This brief was produced with support from Arnold Ventures.

The following organizations support dissemination of MDRC publications and our efforts to communicate with policymakers, practitioners, and others: Arnold Ventures, Ascendium Education Group, Yield Giving/MacKenzie Scott, and earnings from the MDRC Endowment. Contributors to the MDRC Endowment include Alcoa Foundation, The Ambrose Monell Foundation, Anheuser-Busch Foundation, Bristol-Myers Squibb Foundation, Charles Stewart Mott Foundation, Ford Foundation, The George Gund Foundation, The Grable Foundation, The Lizabeth and Frank Newman Charitable Foundation, The New York Times Company Foundation, Jan Nicholson, Paul H. O'Neill Charitable Foundation, John S. Reed, Sandler Foundation, and The Stupski Family Fund, as well as other individual contributors.

For information about MDRC and copies of our publications, see our website: www.mdrc.org.

New York
200 Vesey Street, 23rd Flr.
New York, NY 10281
Tel: 212 532 3200

Washington, DC
750 17th Street, NW
Suite 501
Washington, DC 20006

Oakland
475 14th Street, Suite 750
Oakland, CA 94612
Tel: 510 663 6372