# Understanding and Improving How Policymakers Respond to Program Impact

*Decision aids can help policymakers incorporate evidence in program funding decisions.*

**Target a Priority Outcome** The 2018 Foundations for Evidence-Based Policymaking Act set the stage for an ambitious agenda on the advancement of evidence production and use by the U.S. government, such that evaluation is now considered a critical agency function.[1,2] However, in order for the growing evidence base to effectively impact decisions about which programs to implement, policymakers must be well-equipped to interpret and use evidence. In particular, responsiveness to information about program impact maximizes the possible benefits of government programs in terms of more lives saved, less wasteful spending, and improved well-being for those affected.

**Translate Behavioral Insights** There are many reasons why policymakers may not fully incorporate evidence-based information about a program when deciding whether it is worth funding. In addition to structural barriers to evidence use (evidence availability, or capacity constraints) and preferences about programs that go beyond impact, assessing the total dollar value of a program is complex and cognitively difficult.[3] The Office of Evaluation Sciences (OES) developed a survey experiment to estimate how responsive decision makers are to information about program impact when assessing the value of a program. OES also designed and tested two decision aids to help clarify the mapping between impact and program value. The first decision aid draws on behavioral insights pointing to the benefits of joint evaluations. It presents two alternative programs together on one screen rather than in isolation ("Side-by-Side", see Figure 2).[4] The second decision aid translates total program cost into an annual cost per person impacted based on different features of program impact ("Impact Calculator").

**Embed Evaluation** To better understand whether and when policymakers incorporate evidence-based information about program impact into funding decisions, OES administered a survey experiment among 192 federal employees recruited from across 22 U.S. government agencies from May to October, 2021. 1,470 federal employees with roles that involve developing or interpreting evidence or making program funding decisions were contacted to participate in the survey, with a take up rate of 13%. The median completion time for the survey was 23 minutes.

In the evaluation, respondents reported their maximum willingness to pay for hypothetical government programs (i.e. assessed the value of programs), based on three randomized features relevant to impact: "scope" (number of people reached), "outcome type" (whether the program affects downstream or intermediate outcomes), and "persistence" of effects (how long the program effects last). Assessments were made with and without the two decision aids (i.e. Control, Side-by-Side, and Impact Calculator assessments), in random order. These assessments facilitate the estimation of responsiveness to impact--the extent to which assessments of program value scale in response to a change in program impact--at baseline as well as tests of the efficacy of the two decision aids in improving policymakers' ability to understand and respond to evidence.

**Analyze Data** The evaluation was administered using an online survey platform. The primary outcome of interest is respondents' assessments of program value, which is used to estimate responsiveness to program impact.[5] The evaluation also elicited respondents' certainty in their assessments to explore whether responsiveness is correlated with confidence. Finally, OES collected data on demographics and policymaker experience with evidence and evidence utilization.

[1] 1 Pub. L. No. 115-435, 132 Stat. 5529 (2019), available at https://www.congress.gov/J I5/plaws/publ435/PLA W-I I 5publ435.pdf.
[2] Office of Management and Budget. Evidence-Based Policymaking: Learning Agendas and Annual Evaluation Plans (2021), https://www.whitehouse.gov/wp-content/uploads/2021/06/M-21-27.pdf
[3] Enke, B. and Graeber, T. "Cognitive Uncertainty." Working Paper. (2021).
[4] Hsee, C., G. Loewenstein, S. Blount, and M. Bazerman. "Preference reversals between joint and separate evaluations of options: A review and theoretical analysis." *Psychological Bulletin* 125 no. 5 (1999): 576-590.

[5] Unless noted otherwise, all of the analysis reported in this abstract was prespecified in an analysis plan, which can be found at https://oes.gsa.gov.

**Results** First, the results show that policymakers' responsiveness to program impact is just 0.33; that is, when program impact increases by 100%, policymakers' assessment of the value of the program increases by 33% (90% CI [21%, 45%]). Policymakers are relatively more responsive to persistence (0.59) compared to scope and outcome type (0.24 and 0.23, respectively).

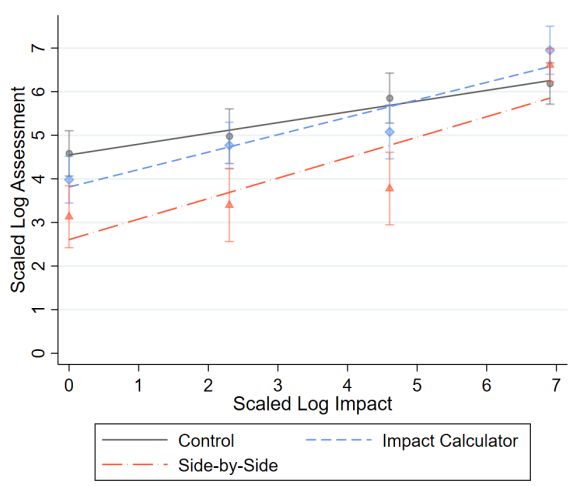


***Figure 1:*** *Treatments increase responsiveness to impact, as reflected by the steeper slopes compared to the Control condition.*

Second, the two decision aids have large and statistically significant impacts on responsiveness. The presentation of two alternative programs Side-by-Side increases responsiveness by 79%, or by 0.26 on the base of 0.33 ($p = 0.001$). The Impact Calculator, which translates total program cost into a cost per person impacted per year, increases responsiveness to program impact by 60%, or by 0.20 on the base of 0.33 ($p = 0.022$).[6]

Finally, correlational evidence shows that responsiveness to impact is related to greater certainty in assessments as well as more experience with the types of program assessments included in the experiment.[7]

Both the efficacy of the decision aids as well as the correlational evidence suggests that the cognitive difficulty of responding to information about impact when assessing a program's value plays an important role in explaining the low responsiveness to impact at baseline.

**Build Evidence** These results identify decision aids that evaluators, policymakers, and researchers can use to more effectively disseminate the results of program evaluations. For instance, written materials may include information on existing programs to give context to a program change (Side-by-Side). They may also highlight information about the number of people a program reaches, program outcomes, and persistence of effects and even combine them with cost data to translate them into a single metric for "cost per unit of impact" (Impact Calculator). Other metrics such as effect sizes could be incorporated in similar ways. The results also suggest that decision makers would be more responsive to program impact if funding decisions were made together on one day or by otherwise comparing multiple proposals together.

Further research might explore the generalizability of these results, specifically the extent to which responsiveness to impact depends on the context in which the decisions are made as well as the role of the respondent in the decision-making process.

***Figure 2:*** *Side-by-Side presentation in the survey experiment.*

Consider two community-based programs that provide person-centered care to people with Alzheimer's Disease and Related Dementias (ADRD).

| Program A | Program B |
|---|---|
| **Number of people reached:** **5 million** individuals with ADRD nationwide. | **Number of people reached:** **5,000** individuals with ADRD in one community. |
| **Program outcome:** In an evaluation researchers found that the program increased the likelihood that an individual with ADRD **is able to live independently** by 18 percentage points. | **Program outcome:** In an evaluation researchers found that the program increased the likelihood that an individual with ADRD **is able to live independently** by 18 percentage points. |
| **How long effects last:** The evaluation lasted 7 years and found that the effects lasted for the first **5 years**. | **How long effects last:** The evaluation lasted 7 years and found that the effects lasted for the first **5 years**. |

Imagine you have the opportunity to help inform how different agencies allocate their budget.

[6] These effects are robust to multiple hypothesis corrections.
[7] The measure used to capture experience was not pre-registered.