

What policymakers want

Documenting a mismatch in policymaker demand and research supply

Max Spohn, Jessica Lasky-Fink, and Elizabeth Linos

The People Lab · July 2026

ABSTRACT

Evidence-based policymaking hinges on policymakers' ability to find, understand, and use evidence in their policy decisions. Building on a growing body of research, we conduct a conjoint experiment to examine how policymakers trade off different dimensions of scientific evidence, including costs, effect sizes, political context, or evaluation methods. We find that policy recommendations are most influenced by program effect sizes as well as contextual and feasibility factors, while measures of internal and external validity (e.g., evaluation method, statistical significance, or sample size) have little impact. We complement our findings with a landscape analysis of nearly 16,000 published research documents and program evaluations showing that there exists a mismatch between the evidence dimensions commonly reported in research and the dimensions that policymakers look for and value most when making decisions.

— FINDING 01

Policymakers' policy recommendations are primarily driven by effect sizes – programs with large effect sizes are 28 pp more likely to be recommended.

— FINDING 02

Policymakers are largely insensitive to internal validity information, but place a lot of weight on contextual factors like political circumstances and implementation.

— FINDING 03

A halo effect in policymakers' perceptions may explain our results – they mistakenly attribute greater internal and external validity to programs with large effect sizes.

— FINDING 04

A demand–supply mismatch: the dimensions policymakers value most – political context, replication, cost – are rarely reported in research.

Introduction

Policymakers are facing growing pressure to incorporate evidence into their decision-making in order to deliver measurable results. But are policymakers able to find – and use – the evidence they need to make decisions? In practice, this is not a straightforward exercise. Rather than simply determining whether an intervention works, policymakers must consider a range of factors that shape how evidence can be interpreted and applied in their context, from effectiveness and methods to implementation costs and feasibility. Yet some of these dimensions may only be partially reported in studies – or absent altogether. As a result, policymakers looking to make evidence-informed decisions must make trade-offs between multiple dimensions of research findings that may not be directly comparable or even observable. We know surprisingly little about how policymakers navigate these trade-offs in practice, even though they are a critical element of the evidence-to-policy pipeline.

This brief examines how policymakers make these judgments. We focus primarily on local government employees who are often responsible for designing, implementing, and managing programs, rather than elected officials, whose use of evidence has been more well-studied (e.g., Baekgaard et al., 2019; Christensen & Moynihan, 2024; Lee, 2022) – though our sample includes both (see Table 1). Using a series of conjoint experiments, we study how local government employees across the U.S. weigh different dimensions of evidence when making recommendations about scaling a policy. We analyze which dimensions carry the most weight, how they interact in decision-making, and how patterns vary across policymakers. We also test whether alternative framings of effect sizes and costs influence decisions. Finally, we complement the experimental results with an LLM-assisted analysis of policy research documents to assess whether the information policymakers value aligns with what is typically reported.

What we did

In November 2025, we conducted a survey experiment among 1,713 local policymakers across the U.S. In a conjoint design (Hainmueller et al., 2014), participants were presented with a series of tables that summarized two hypothetical program evaluations side-by-side. All programs in the experiment addressed challenges in housing, criminal justice, or employee wellbeing. The tables included information on 15 different evidence dimensions for each program and the levels of each dimension were randomized. For example, the program’s evaluation method was described either as a focus group, correlational data, before-and-after comparison, or a randomized experiment. Table 1 shows an overview of the evidence dimensions and example levels; the Appendix shows additional details.¹

After reviewing each set of tables, participants were asked which, if any, of the two programs they would recommend for pilot testing or implementation. This approach allows us to isolate the causal impact of each dimension on whether a program is recommended: for instance, does increasing the sample size of an evaluation from 100 to 1,000 participants make a program more likely to be recommended for implementation? Does reporting that a program is costly make it more likely to be recommended for testing, relative to not mentioning cost at all? By varying these features systematically, we can disentangle how policymakers weigh and trade off different dimensions of evidence in a complex information environment. This work builds on prior literature that suggests that contextual factors, such as local relevance or implementer expertise, can matter as much as traditional markers of rigor and impact – but that largely examines these dimensions in isolation (Hjort et al., 2021; Nakajima, 2021; Vivalt & Coville, 2025).²

In addition, we measured participants’ perceptions of programs’ effectiveness, cost, internal validity, external validity, and feasibility. We also collected demographic and background data, including experience with policy evaluations.

Table 1. Overview of evidence dimensions in the conjoint experiment

DIMENSION	DESCRIPTION
Type of program	Brief description of the program (e.g., “Free legal counsel in eviction court”)
Theory of change	Brief explanation of why and how the program is expected to affect outcomes
Length of program	Program duration (e.g., “One month”)
Evaluation method	Brief description of the evaluation method: focus group reports / correlational data / before-and-after data / randomized experiment
Sample size	Sample size in the evaluation (e.g., “~1,000 renters”)
Location	Location of the evaluation (e.g., “In a neighboring city in the same state”)
Year	Year of the evaluation: 1981 - 2022
Immediate program outcome	Brief description of the result: null / backfire / small effect / large effect <i>Results were described in narrative terms (e.g., “slightly higher”), as a percentage point change, as a percentage change, as a difference from the baseline, or as a standardized effect size (SD units).</i>
Statistical significance	Statistical significance of the result: not reported / significant / insignificant <i>Statistical significance was described in narrative terms (e.g., “the result was statistically significant”), with p-values, with standard errors, or with 95% confidence intervals.</i>
Subgroup differences	Brief description of subgroup effects: not reported / no difference / worse / better <i>Subgroups included Black participants and context-specific vulnerable populations – the descriptions of “worse” and “better” were phrased to correspond to the main results.</i>
Long-term program outcome	Brief description of results after 2 years: not reported / no difference / worse / better <i>The descriptions of “worse” and “better” were phrased to correspond to the main results.</i>
Cost of the program	Brief description of the program costs: not reported / low / moderate / high <i>Costs were described in narrative terms (e.g., “low cost”), as per-participant costs, as total costs, as a cost-benefit ratio, or as net \$ saved per \$ spent.</i>
Study replications	Brief description of direct replications: not reported / no difference / worse / better <i>The descriptions of worse and better were phrased to correspond to the main results</i>
Existing processes	Brief description of existing processes relevant to implementation (e.g., “CITYVILLE is already organizing informational workshops for residents that can be adapted.”)
Political context	Brief description of political context relevant to implementation (e.g., “CITYVILLE’s councilmembers expressed support for the program.”)

¹ In a series of pilots, we tested whether varying the number of dimensions affects participants’ choices but found no systematic differences between tables with fewer than 10 dimensions and tables with more than 20 dimensions.

² Respondents were randomly assigned to one of two outcomes: (1) which program would you recommend for implementation, or (2) which program would you recommend for pilot testing? Results do not differ by outcome. Thus, all findings reported here include all respondents, regardless of which outcome they saw, and control for experimental condition.

Sample

Our sample was drawn from 227,721 current local government employees across 3,208 U.S. cities and 2,405 U.S. counties with more than 10,000 inhabitants. Using publicly available domain information, we predicted their email addresses and disseminated the study via email. In total, 1,713 policymakers (0.7%) responded, covering 48 U.S. states and a wide range of roles ranging from elected officials to heads of departments and staff employees. Our sample is somewhat skewed towards more highly educated respondents and self-identified Democrats compared to national estimates of the state and local workforce (U.S. Bureau of Labor Statistics, 2022). Table 2 provides an overview of the sample characteristics.

Table 2. Overview of sample characteristics · N = 1,713

EMPLOYER TYPE

City employees	62.4%
County employees	37.6%

AGE

Average age	44 years
-------------	----------

GENDER

Female	55.7%
Male	41.9%

RACE

White	83.3%
Black	6.3%
Other	10.4%

POLITICAL PARTY

Democrat (incl. leaners)	71.5%
Republican (incl. leaners)	16.9%

ROLE

Publicly elected official	2.5%
City manager / administrator	2.0%
Head of dept. / chief of staff	10.7%
Lower-level manager	38.5%
Frontline worker / staff	38.6%

DEPARTMENT

Housing / public safety / HR	24.7%
Other departments	75.3%

EDUCATION

High school / equivalent	1.6%
Some college / associate's	10.3%
Bachelor's degree	37.4%
Master's degree	41.5%
Doctorate / professional	8.9%

What we found

Policymakers value effect sizes more than any other evidence dimension.

Effect size is the single most influential dimension shaping policymakers' policy recommendations. Policymakers are 28 percentage points (pp) more likely to recommend a program with a large effect than one with a null effect (Figure 1) – more than three times the impact of any other attribute we test. They are also 6 pp more likely to recommend programs with long-term effects that exceed those short-term effects. We also find evidence that the framing of effect sizes matters: programs with large effect sizes are more than 10 pp more likely to be recommended when they are framed as percentage changes rather than standardized effect sizes or percentage point changes (Figure 2), underscoring how seemingly technical choices in reporting may meaningfully shape policy decisions.

Importantly, the weight policymakers place on large effect sizes is not conditional on evaluation method (Figure 3a) – that is, all evaluation methods are treated as equally credible when determining “what works.” The weight placed on effect sizes is also not conditional on program cost (Figure 3b) or any other evidence dimension. This suggests that effect size serves as a dominant signal in decision-making, even when other relevant information is available.

One possible reason for the outsized influence of effect sizes on decision-making appears to be a halo effect. Policymakers infer broader program quality from large effects: holding all else constant, they perceive programs with large effect sizes as more internally valid (Cohen's $d = 0.33$) and more externally valid ($d = 0.17$) than programs with null results (Figure 4). In other words, large effects lead policymakers to make additional – and potentially unwarranted – inferences about the credibility and generalizability of the evidence.

HALO EFFECT

Large effects make a program seem more internally valid – even when nothing else changes.

These inferences stand in contrast to other dimensions, where perceptions are generally well-calibrated. For example, policymakers correctly rate high-cost programs as more expensive than low-cost ones ($d = 0.38$), and programs built on existing processes as more feasible ($d = 0.15$). Taken together, these results suggest that policymakers may overweight effect sizes not only directly, but also indirectly – by allowing large impacts to shape their perceptions of other evidence dimensions, potentially crowding out attention to methodological rigor.

Figure 1. Effects of main evidence dimensions on program recommendation.

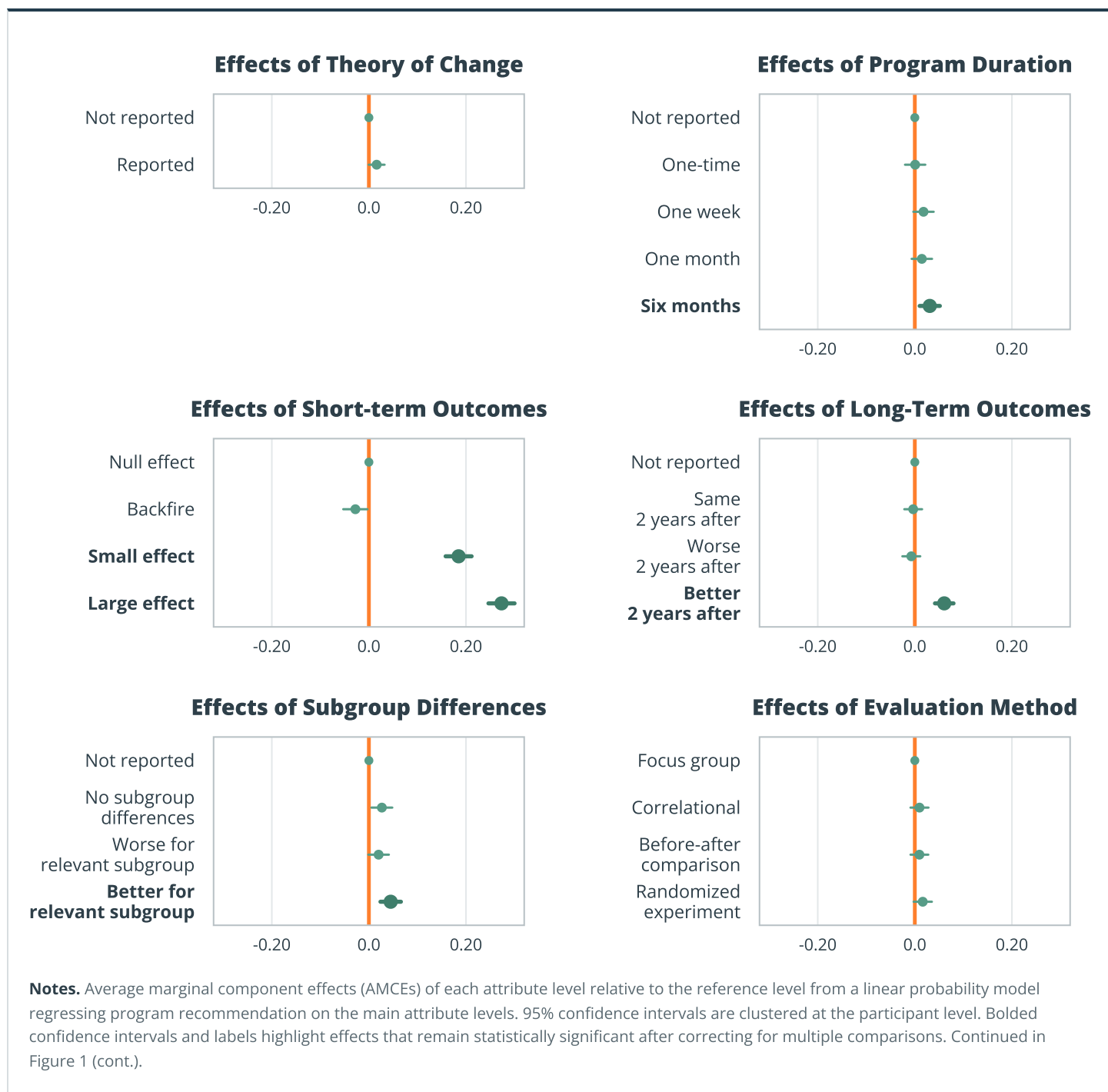


Figure 1 (cont.). Effects of main evidence dimensions on program recommendation.

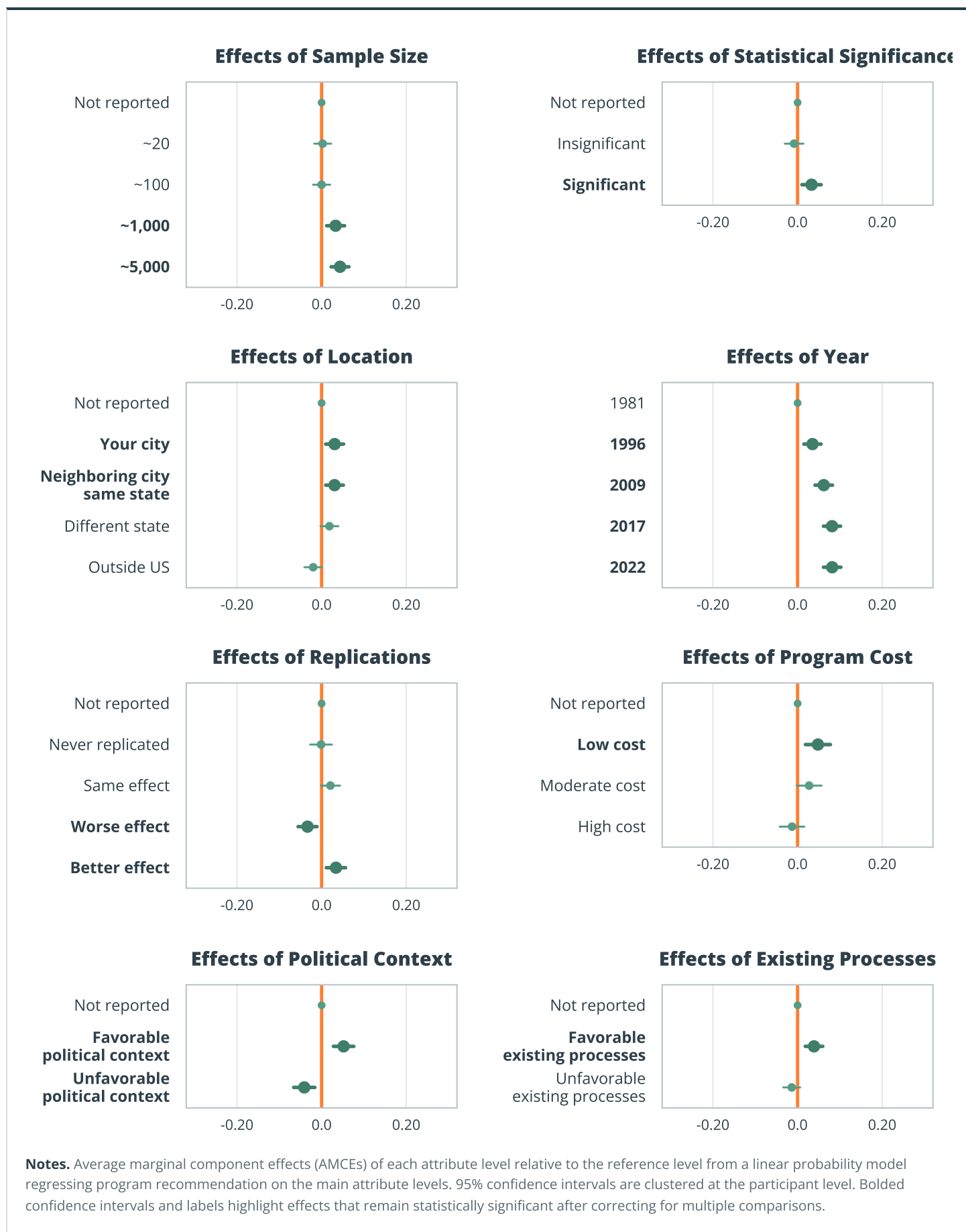


Figure 2. Effects of short-term outcome framings on program recommendation.

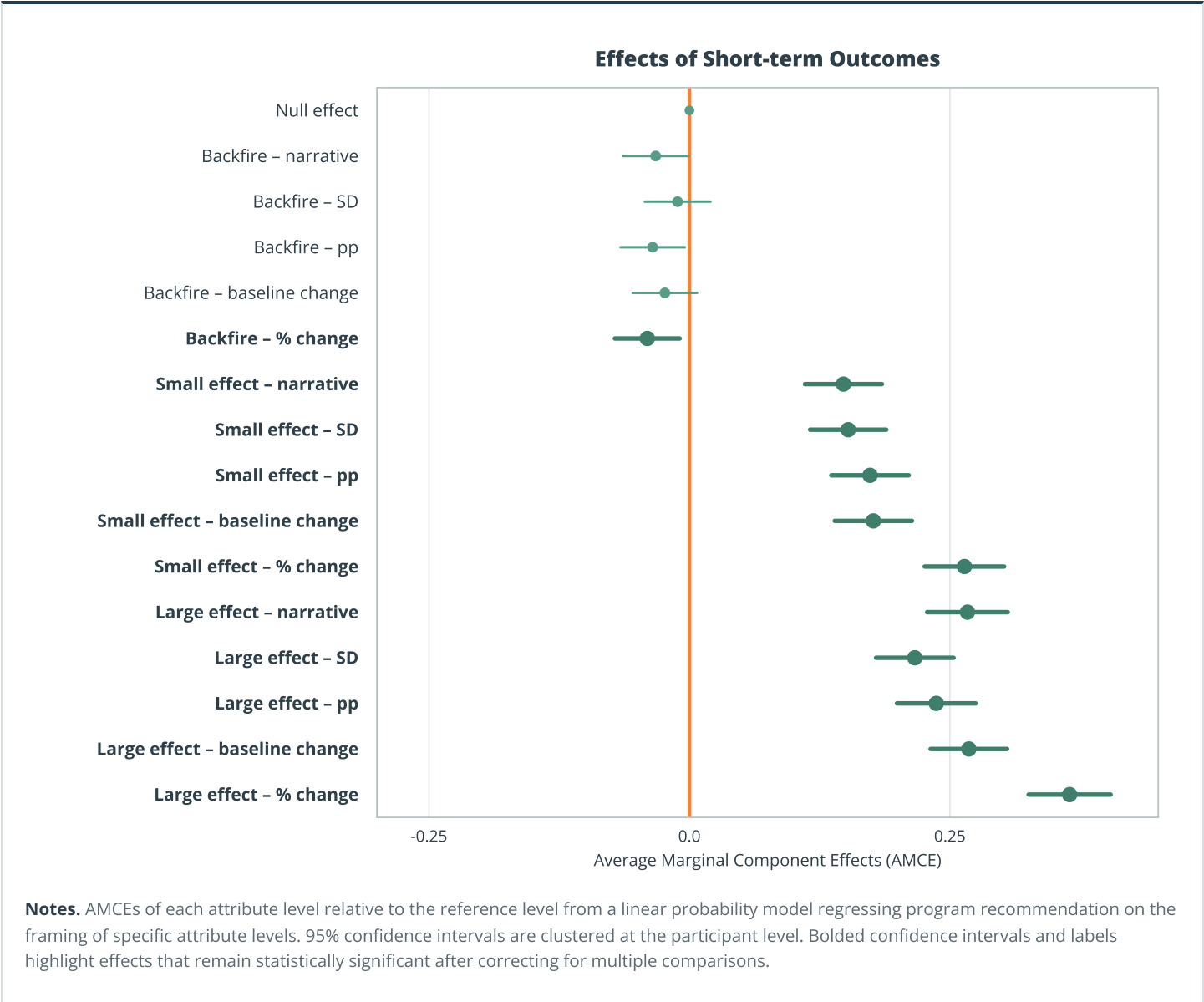


Figure 3. Effects of short-term outcomes, conditional on evaluation method and program cost.

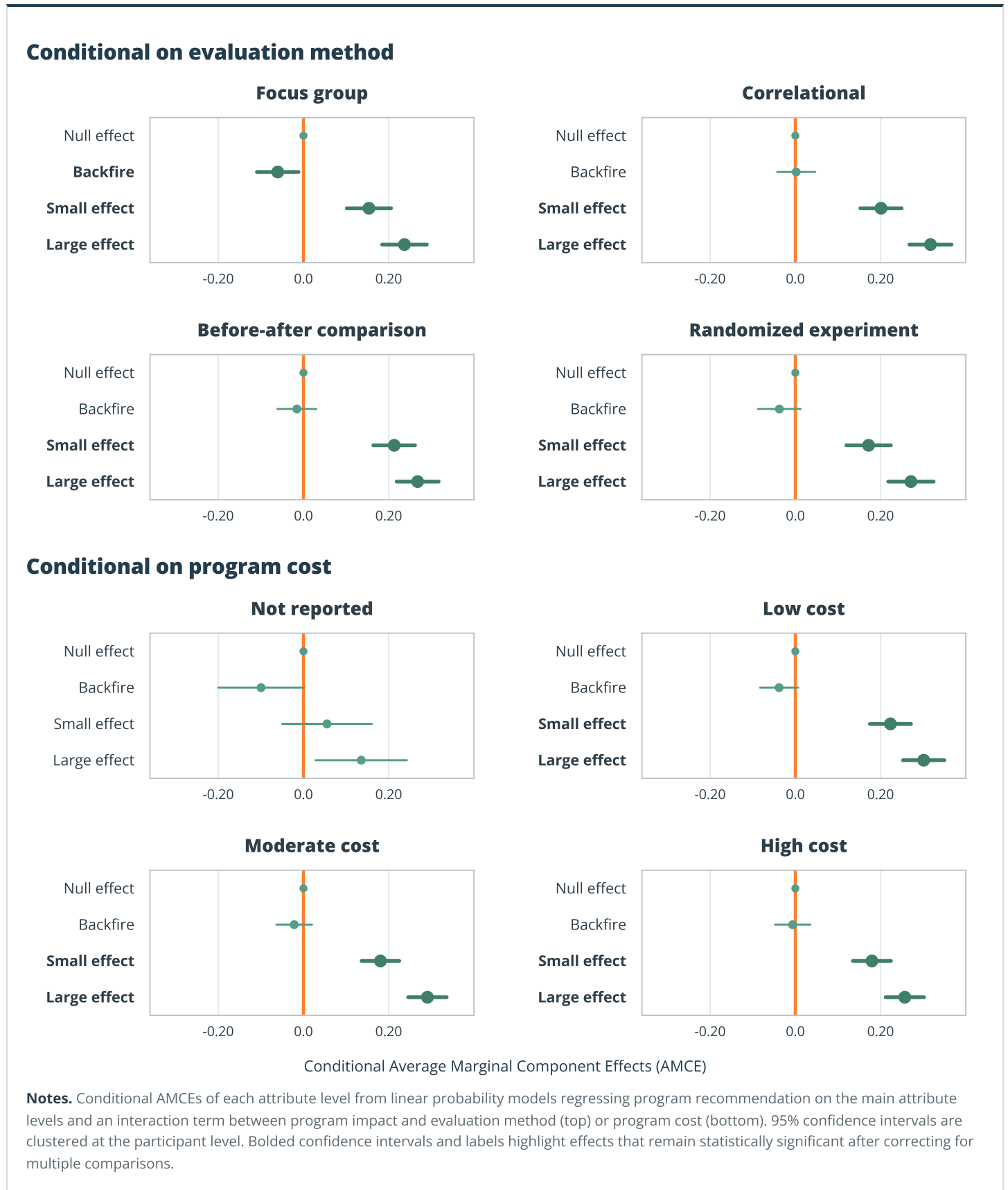
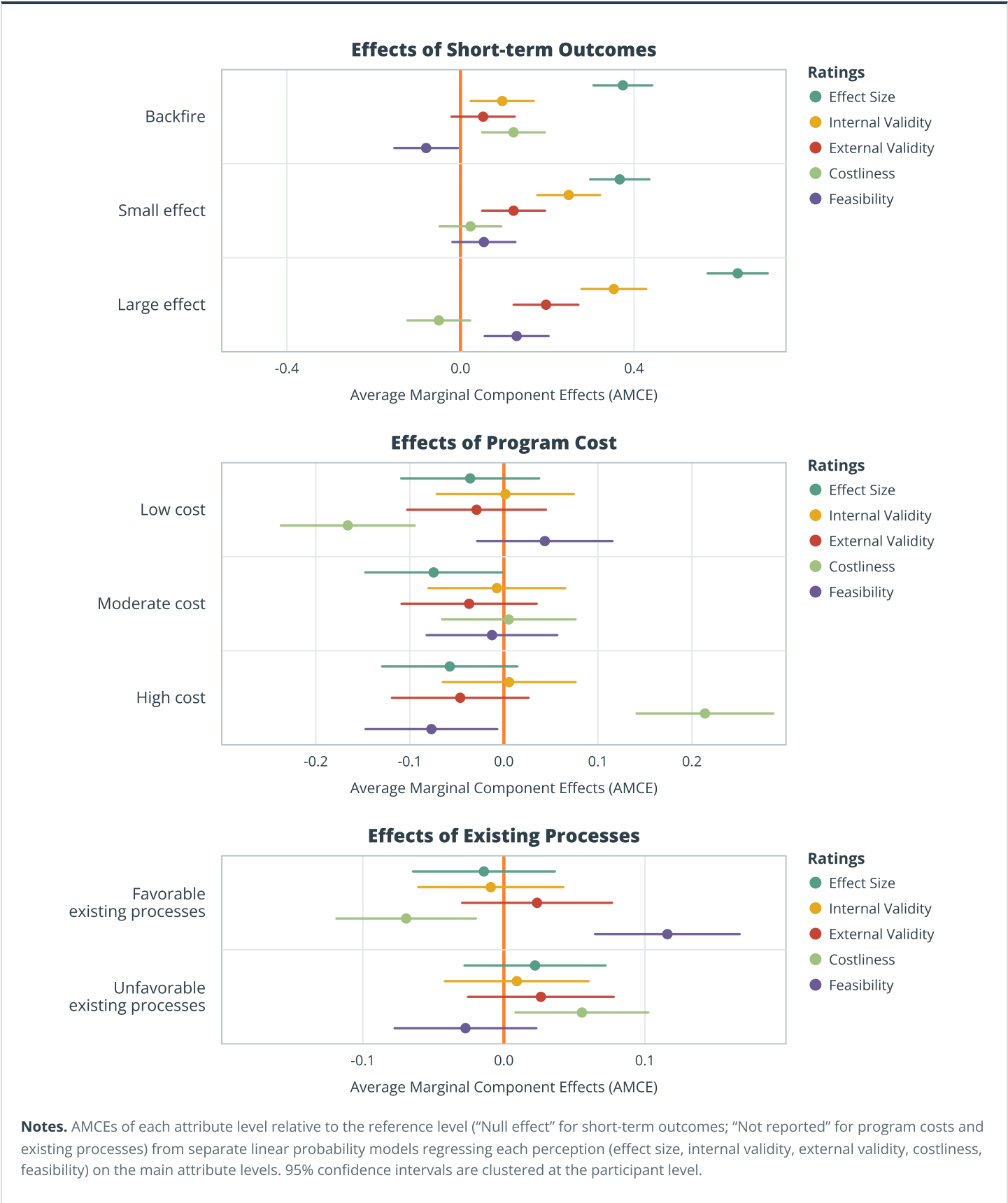


Figure 4. Effects of short-term outcomes, program costs, and existing processes on perceptions of effect size, internal validity, external validity, costliness, and feasibility.



Policymakers are relatively insensitive to internal and external validity measures.

Consistent with prior research (e.g., Nakajima, 2021; Xu et al., 2025), traditional markers of internal validity play a comparatively small role in policymakers' decisions. In particular, evaluation methods have little influence: programs evaluated using randomized experiments are only 2.2 pp more likely to be recommended than those based on focus groups – a difference that is not statistically significant (Figure 1). In fact, a program's evaluation method emerges as one of the least influential dimensions in our experiment.

Policymakers also appear relatively insensitive to other indicators of internal validity. Increasing the sample size from 20 to 5,000 participants raises the likelihood of recommendation by just 3.2 pp, and statistically significant results increase recommendations by only 4.2 pp relative to statistically insignificant results.

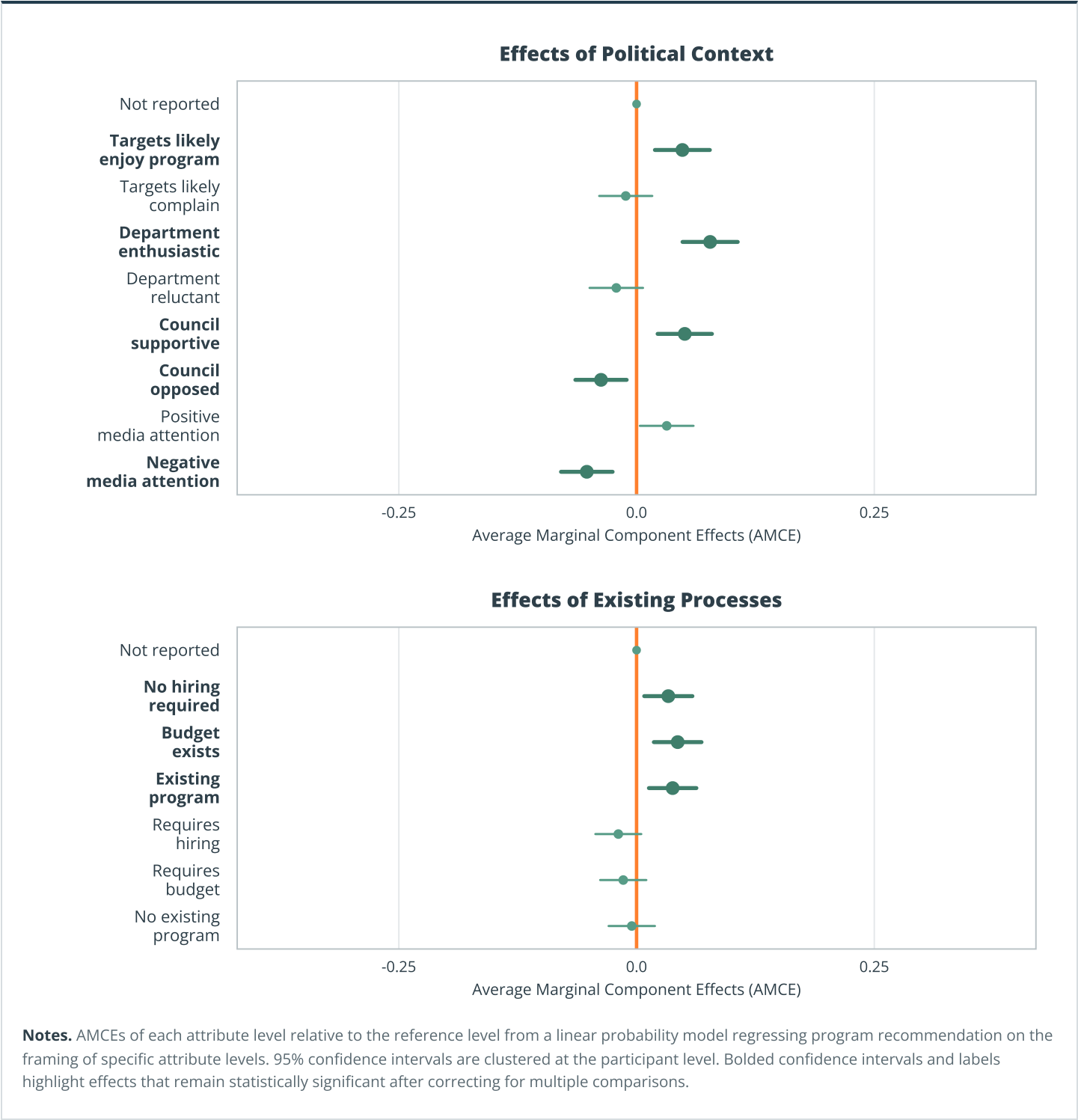
Similarly, evidence of successful replication – which could be a measure of either internal or external validity – does not meaningfully affect recommendations unless the replication produces larger effect sizes than the original study. Other measures of external validity, such as location and year, also have relatively small effects on program recommendation.

Context and feasibility indicators meaningfully shape decisions.

Policymakers are responsive to signals about political support and implementation feasibility, though these factors carry less weight than effect sizes. Programs implemented under favorable political conditions are 8.9 pp more likely to be recommended than those facing unfavorable conditions (Figure 5, top panel). Indicators such as expected city council support, enthusiasm from implementing departments, and positive reactions from target populations all increase the likelihood of recommendation, while negative media attention reduces it.

Feasibility within existing systems also matters (e.g., DellaVigna et al., 2024). Policymakers are 5.4 pp more likely to recommend programs that can be implemented using existing processes – for example, those that do not require additional hiring, fit within current budgets, or build on existing programs (Figure 5, bottom panel). These findings suggest that practical considerations around implementation meaningfully shape decision-making alongside reported impact.

Figure 5. Effects of political context and existing processes on program recommendation.



Costs matter – but framing has limited impact.

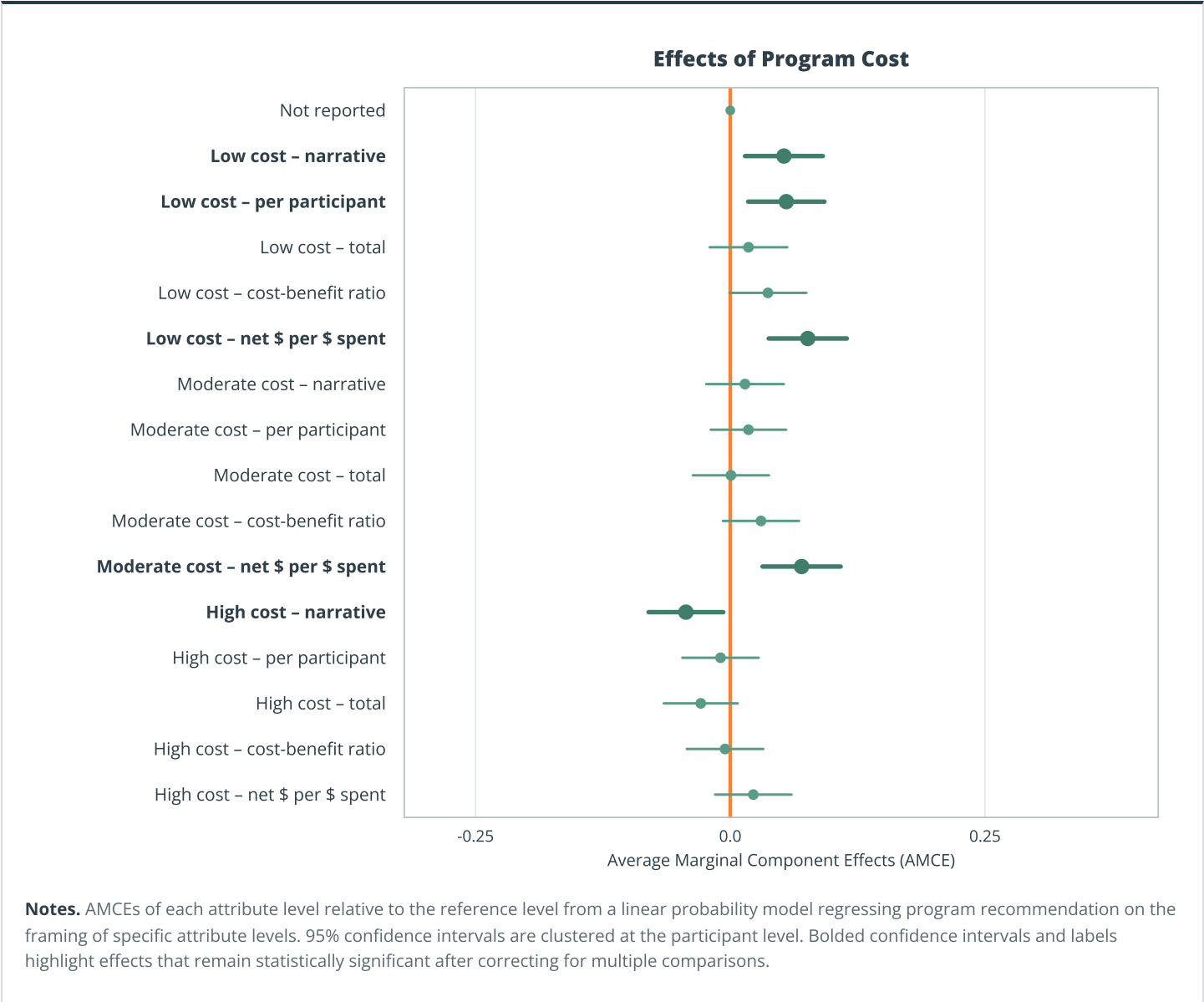
Costs impact policymakers' recommendations, but relatively modestly: regardless of the policy area, programs with lower costs are 5.7 pp more likely to be recommended than higher-cost alternatives. In contrast to effect sizes, we find little evidence that how costs are framed substantially affects decisions. Policymakers are only slightly more likely to recommend programs when costs are presented as dollars saved per dollar spent, regardless of the underlying cost (Figure 6). Cost-benefit calculations may help policymakers make sense of financial information that would otherwise be difficult to evaluate, correcting somewhat distorted reactions to high program costs. Overall, however, cost considerations appear secondary to impact considerations, and less sensitive to presentation.

Little heterogeneity in policymakers' recommendations.

We find surprisingly little heterogeneity in policymakers' recommendations, though we are underpowered for many subgroup analyses. For instance, we find no differences in the effect of different evidence dimensions on policymakers' recommendations by policy domain (housing, criminal justice, employee wellbeing). Relatedly, we find no meaningful heterogeneity by the policy area in which the respondent works, nor by their role (e.g., supervisory role vs. frontline worker).

Policymakers with graduate degrees appear to place somewhat more weight on statistical significance, while those who have higher trust in scientists, greater experience with scientific research, or better understanding of the evidence dimensions presented in the conjoint tables are more sensitive to evaluation methods, especially randomized evaluations. While these differences are substantively meaningful, almost all fail to reach conventional levels of statistical significance after adjusting for multiple comparisons.

Figure 6. Effects of program cost framings on program recommendation.



A mismatch between policymakers and researchers?

Our conjoint experiment shows that policymakers prioritize some dimensions of evidence far more than others. They place substantial weight on effect sizes and on contextual and implementation factors, such as costs, political support, and whether programs can build on existing processes. At the same time, they are relatively insensitive to information that typically signals the credibility of results, including evaluation methods and sample sizes. This pattern raises an important question: does the evidence available to policymakers align with what they value?

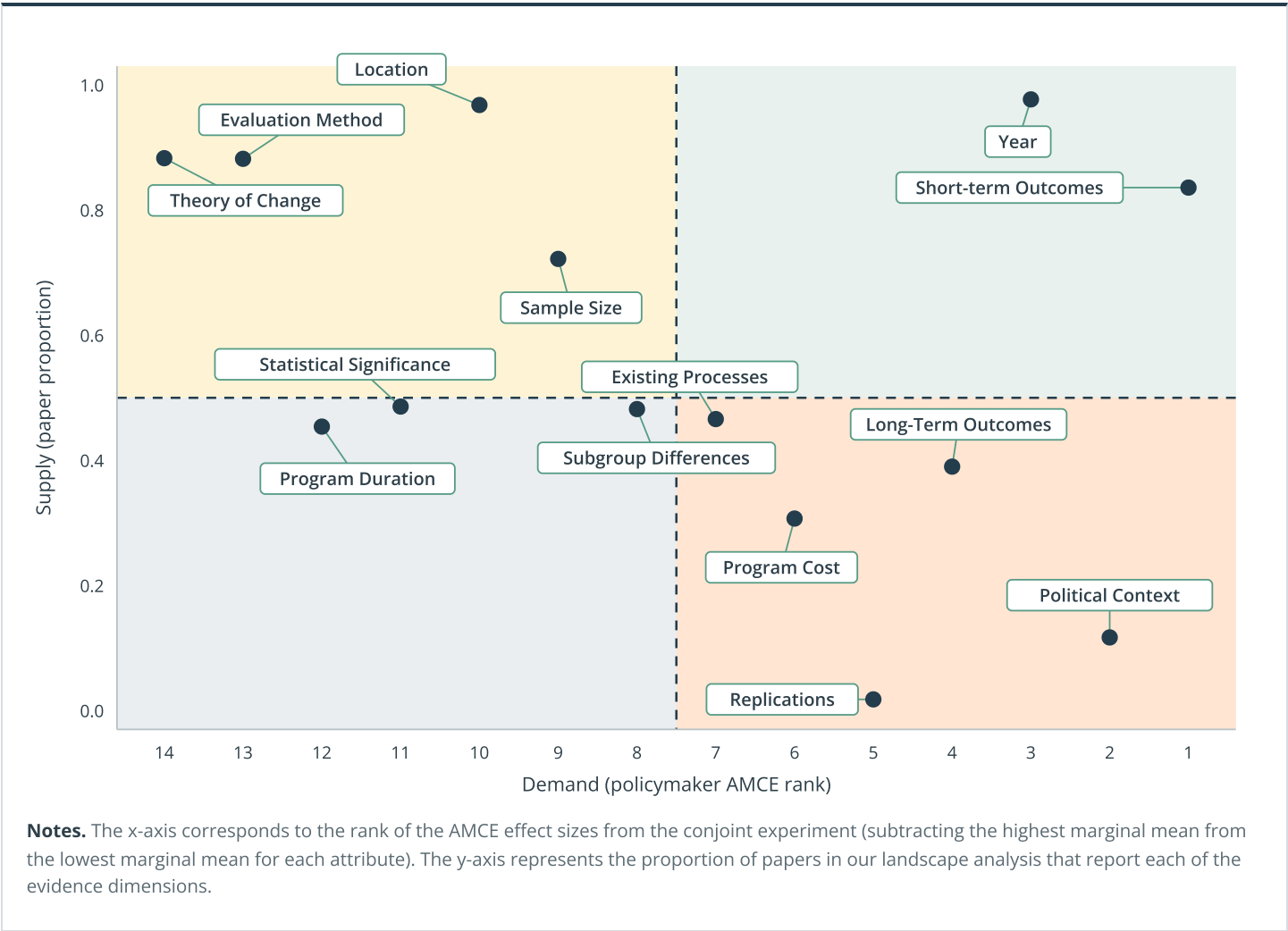
To answer this, we draw on a landscape analysis (see [*What makes research adoptable at scale? A landscape analysis of policy research*](#)) in which we examine how evidence is reported in policy research. Using GPT-4o, we coded 15,995 documents from 27 evidence clearinghouses, spanning academic articles, research reports, policy briefs, and white papers. For each document, we recorded whether it includes the same evidence dimensions presented in our conjoint experiment, allowing us to estimate how frequently each dimension is reported. This approach comes with a series of limitations: LLM-coded features of research documents may be noisy, unreliable, or systematically biased. Our focus on binary codes for these features may also miss important heterogeneity in how much information and depth different documents provide on each evidence dimension. Nevertheless, our approach allows us to scale the landscape analysis to thousands of documents that become comparable through our coding scheme, and may mimic the level of depth and time that a given policymaker has to scan an academic study.

Comparing the results of our landscape analysis and conjoint experiment reveals a clear mismatch between policymakers and researchers (Figure 7). On one hand, policymakers value several dimensions that are rarely reported in research documents. Political context – one of the most influential factors in our experiment – is mentioned in fewer than 12% of documents. Similarly, fewer than 2% of documents report on replication, and only about 30% include information on program costs, despite the importance of these evidence dimensions for policymaker decision-making.

On the other hand, policymakers place relatively little weight on dimensions that researchers emphasize heavily. Evaluation methods and theory of change are each reported in roughly 88% of documents, while 72% report sample sizes and 49% report statistical significance. Among studies using causal identification strategies (e.g., RCTs, RDDs, DiD, IV), these shares are even higher: 98% report a theory of change, and 93% report both sample sizes and statistical significance. Researchers' emphasis on methodological rigor in identifying what works is much less appreciated by policymakers.

Our evidence suggests that there is a systematic misalignment between what policymakers value, and how evidence is commonly presented. Researchers consistently report dimensions that policymakers largely discount, while omitting information that policymakers find critical for decision-making. Closing this gap will be essential for improving the uptake and use of evidence in practice.

Figure 7. Supply of and demand for evidence dimensions in policy research.



What's next

Overall, our findings suggest that one barrier to evidence adoption at scale may be a misalignment between what policymakers value for decision-making and how research is presented. We find that policymakers' decisions are driven largely by effect sizes, regardless of the quality or credibility of the underlying evidence. At the same time, they also value contextual and implementation factors, such as political support and feasibility, that are rarely emphasized in academic research.

While many research documents – especially academic papers – may not be explicitly aimed at policymakers, they are still one of the main conduits of information in the evidence-to-policy pipeline. A recent survey among Australian policymakers found that more than 90% of policymakers search the internet for research, yet less than a third email or phone academics about their work (Head et al., 2014). Moreover, even evidence clearinghouses and other tools that synthesize available evidence are typically limited by what information is included in original papers: overall, 34% of documents in our landscape analysis are academic papers; 53% are policy reports. Thus, if researchers don't report contextual or feasibility information, that crucial information may never reach policymakers. This gap creates a disconnect in the evidence-to-policy pipeline that undermines the promise of evidence-based policymaking.

From a practical perspective, closing this gap will require a bi-directional approach. On the one hand, policymakers may benefit from support in translating evidence – through training, tools, or decision aids that help distinguish credible from non-credible findings (Mehmood et al., 2023). On the other hand, researchers may need to adapt how they produce and communicate evidence. This could include standardizing the presentation of key evidence dimensions (e.g., Toma & Bell, 2024), as well as systematically reporting contextual factors that are central to real-world decision-making. Future work should test interventions that align how evidence is produced, communicated, and used – bringing the promise of evidence-based policymaking closer to practice.

ABOUT TPL

The People Lab aims to empower the public sector by producing cutting-edge research on the people of government and the communities they serve. Using evidence from public management and insights from behavioral science, we study, design, and test strategies for solving urgent public sector challenges in three core areas: strengthening the government workforce; improving resident–government interactions; and reimagining the production and use of evidence.

Acknowledgements: This research was supported by Arnold Ventures.



The People Lab



HARVARD Kennedy School
JOHN F. KENNEDY SCHOOL OF GOVERNMENT

CONNECT WITH US



peoplelab.hks.harvard.edu



LinkedIn



Bluesky

References

- Baekgaard, M., Christensen, J., Dahlmann, C. M., Mathiasen, A., & Petersen, N. B. G. (2019). The role of evidence in politics: Motivated reasoning and persuasion among politicians. *British Journal of Political Science*, 49(3), 1117–1140.
- Christensen, J., & Moynihan, D. P. (2024). Motivated reasoning and policy information: Politicians are more resistant to debiasing interventions than the general public. *Behavioural Public Policy*, 8(1), 47–68.
- DellaVigna, S., Kim, W., & Linos, E. (2024). Bottlenecks for evidence adoption. *Journal of Political Economy*, 132(8), 2748–2789.
- Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2014). Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis*, 22(1), 1–30.
- Head, B., Ferguson, M., Cherney, A., & Boreham, P. (2014). Are policy-makers interested in social research? Exploring the sources and uses of valued information among public servants in Australia. *Policy and Society*, 33(2), 89–101.
- Hjort, J., Moreira, D., Rao, G., & Santini, J. F. (2021). How research affects policy: Experimental evidence from 2,150 Brazilian municipalities. *American Economic Review*, 111(5), 1442–1480.
- Lee, N. (2022). Do policy makers listen to experts? Evidence from a national survey of local and state policy makers. *American Political Science Review*, 116(2), 677–688.
- Mehmood, S., Naseer, S., & Chen, D. L. (2023). Training policymakers in econometrics. Presented at the ASSA Annual Meeting, January 6–8, 2023, New Orleans, LA.
- Nakajima, N. (2021). Evidence-based decisions and education policymakers. Unpublished paper.
- Toma, M., & Bell, E. (2024). Understanding and increasing policymakers' sensitivity to program impact. *Journal of Public Economics*, 234, 105096.
- U.S. Bureau of Labor Statistics. (2022, Jan 3). State and local government jobs more likely than private sector to require postsecondary education.
- Vivalt, E., & Coville, A. (2025). Local knowledge, formal evidence, and policy decisions. *Journal of Development Economics*, 174, 103425.
- Xu, C., Cheng, Y., Wang, S., Merrick, W., & Carter, P. (2025). Evaluating use of evidence in US state governments: A conjoint analysis. *Public Administration Review*, 85(4), 1217–1235.

Study Design

Participants were randomly assigned to one of three policy areas: employee burnout, housing, or public safety. Then, in a conjoint design, all participants saw four tasks – each task presented a table comparing two programs related to the randomly assigned policy area (e.g., two programs aimed at reducing employee burnout, reducing evictions, or reducing youth crime). The tables included 15 distinct evidence dimensions, each with randomly assigned levels, as shown in Table A1.

At the beginning of the survey, participants were shown the instructions to contextualize the experiment. For example, participants assigned to the housing policy area were shown the following instructions:

HOUSING SCENARIO • SHOWN TO PARTICIPANTS

Please imagine the following scenario: You are a public servant working for the housing authority of the fictional city CITYVILLE (USA). You have been asked by your manager to identify promising practices that could be implemented in CITYVILLE to reduce eviction rates among low-income families.

On each of the next four screens you will see tables that contain short summaries of the results of two programs that aimed to reduce eviction rates. Every table will be different. Please read each of the tables carefully to answer the questions that follow.

The situation in CITYVILLE: The current eviction rate in CITYVILLE is 6%, well above the national average of 4% for cities of its size. The standard deviation of eviction rates is 3%, which means that around two-thirds of these cities have eviction rates between 1% and 7%. Reducing the eviction rate would not only help families in need – it could also save the city a lot of money. A homeless shelter bed for an evicted person costs CITYVILLE \$5,000 for 3 months – that’s how long it typically takes evicted people to find new housing in CITYVILLE.

Table A1. Evidence dimensions and levels (policy area: housing)

ATTRIBUTE	LEVELS	EXAMPLE
Type of program	Famous	Information mailers on emergency rental assistance
	Well-known	Free legal counsel in eviction court
	Emerging	Tenant education programs informing tenants about their rights
	Unknown	Mediation between landlords and tenants before eviction filings
Theory of change	Not reported	
	Reported	Legal aid programs supply trained attorneys and caseworkers who represent tenants during eviction hearings.
Length of program	Not reported	
	One-time	
	One week	
	One month	
	Six months	
Evaluation method	Focus group	A guided discussion to collect eviction rates, opinions, and experiences with the program among all participants
	Correlational data	Eviction rates among renters who decided to participate were compared to those who decided not to participate
	Before-and-after	Eviction rates before the program was implemented were compared to rates after, among renters who participated
	Randomized experiment	Eviction rates among renters randomly selected to participate were compared to similar renters not selected
Sample size	Not reported	
	~20 renters	
	~100 renters	
	~1,000 renters	
	~5,000 renters	
Location	Not reported	
	In your city (CITYVILLE)	
	Neighboring city, same state	
	Different state	
	Outside the US	

ATTRIBUTE	LEVELS	EXAMPLE
Year	1981	
	1996	
	2009	
	2017	
	2022	
Immediate program outcome	Null	The eviction rate among program participants did not differ
	Backfire	The eviction rate among program participants was 0.3 pp higher
	Small effect	The eviction rate among program participants was 5.7% compared to the 6% baseline
	Large effect	The eviction rate among program participants was 0.5 standard deviations lower
Statistical significance	Not reported	
	Significant	The p-value associated with the difference in the eviction rate was $p = 0.01$
	Insignificant	With 95% confidence, the true difference lies between -3.95 and 0.95 pp
Subgroup differences	Not reported	
	No difference	The difference in the eviction rate was the same for Black and White participants
	Subgroup worse	There was no difference in the eviction rate among Black participants
	Subgroup better	The eviction rate among participants with prior evictions was lower than among those without
Long-term outcome	Not reported	
	No difference	The eviction rate among participants was the same 2 years after the program ended
	Worse	The difference among participants disappeared 2 years after the program ended
	Better	The eviction rate among participants was even lower 2 years after the program ended
Cost of the program	Not reported	
	Low	\$20 per participant
	Moderate	\$2,000 in total
	High	Benefit-cost ratio: 3:4

ATTRIBUTE	LEVELS	EXAMPLE
Study replications	Not reported	The program was never replicated
	No difference	Replicated in the same city; the difference in the eviction rate was the same as in the original
	Worse	Replicated in a different city; there was no difference in the eviction rate among participants
	Better	Replicated in a different city; the eviction rate was lower than in the original evaluation
Existing processes	No guidance	The evaluation provides no guidance on leveraging existing processes
	Favorable	CITYVILLE is already sending informational mailers to residents that can be adapted
	Unfavorable	The program would require CITYVILLE to hire additional staff
Political context	No guidance	The evaluation provides no guidance on the political context
	Favorable	CITYVILLE's councilmembers expressed support for the program
	Unfavorable	CITYVILLE's housing authority is reluctant to implement the program