INSIGHTS

POLICY FORUM

EVIDENCE-BASED POLICY

Policy evaluation and the causal analysis of public support

Support for initially unpopular policies may grow over time

By **Stefano Carattini**[1], **Robert Dur**[2], **John List**[3]

Many policies that are generally considered socially desirable by the scientific community, based on modeling and causal empirical analyses, are not very widespread. The main driver is often lack of public support at baseline ("ex ante"). Yet, there is evidence that when voters hold bi-ased beliefs ex ante about a given policy, experiencing the policy first-hand may lead them to correct their beliefs and increase public support (1). If it was widely documented that opposition to sound policies in part dissipates when voters experience a given policy, then more policy-makers may be inclined to experiment with policies that scientists recommend but that are unpopular ex ante. Systematically combin-ing policy evaluation with causal analysis of public support would allow scholars to create a body of knowledge on the conditions under which policies become more (or less) popular after implementation and what are the drivers of changes in beliefs and public support.

Consider congestion charges, which make accessing a city center costly. That only a handful of cities around the world have congestion charges in place reflects how little things have changed since 1963, when economist (and later Nobel laureate) William S. Vickrey wrote that "in no other major area are pricing practices so irra-tional, so out of date, and so conducive to waste as in urban transportation" [(2), p. 452]. Or consider carbon taxes. Voters may expect carbon taxes not to change people's behavior and lead to better environmental outcomes, underestimating environmental

To reduce traffic, Stockholm introduced a congestion charge, initially with a trial period so that voters could experience it first-hand.

recent arguments concerning the science of using science: Evidence showing policies becoming more popular ex post could prevent repeals in front-runner jurisdictions, while encouraging policy-makers elsewhere to follow suit. If public support would decrease after implementation, it would be equally important to know.

THE RATIONALE

Voters have been shown to hold biased beliefs on a variety of policies (4). Evidence from laboratory studies shows that people often oppose policies that would make them better off, for lack of understanding (5, 6). However, the same studies show that allowing people to experience the policy contributes to address their biased beliefs and improve demand for policies that make everybody better off.

Laboratory experiments generally provide an ideal context for experimentation, thus having high "internal validity." However, laboratory experiments by design simplify the real world to examine specific questions in an artificial environment. There are, however, several observational studies using field data providing anecdotal evidence suggesting that public acceptability improves once voters experience a new policy. Public support for congestion charges seems to have improved after trial periods in the Swedish cities of Stockholm and Gothenburg (7, 8). The carbon tax that British Columbia introduced in 2008 became more popular over time (9). However, economic conditions also improved globally in the aftermath of the Great Recession, potentially making climate policy more attractive and thus showing the limits of pre-post comparisons for causal interpretation. Hence, a proper counterfactual is necessary also for the causal analysis of public support, as it is for policy evaluation. Policy evaluation and causal analysis of public support would go hand in hand.

To the best of our knowledge, only one study to date combines policy evaluation and the causal analysis of public support (10). This study examines the implementation of pricing garbage by the bag, an environmental tax increasing the cost of residential garbage to encourage recycling. The study combines survey and administrative data to show that pricing garbage by the bag reduces residential garbage by about 40%, while generating very limited illicit dumping and leakage to neighboring communities without the policy. But survey data also documented substantial

opposition to the policy before it was implemented, with many voters considering it ineffective and unfair. Once the policy was implemented, though, voters' perceptions substantially improved. Voters realized that the policy was effective, which also reduced demand for environmental earmarking, a design feature that policy-makers often favor when voters perceive an environmental tax as ineffective (1), further motivating the importance of the causal analysis of public support. An important feature of this study is that it included an already-treated control group, which allowed researchers to measure perceptions among voters who already knew the policy. When the policy expanded to other areas, perceptions among voters who had already experienced the policy were unchanged.

Although here we argue for more evidence on the conditions that make ex ante unpopular policies more popular ex post, we already know from the above-mentioned cases that voters are more likely to see the policy more favorably when they realize that it is (more) effective (than they thought) and has fewer negative consequences. In the case of Swedish cities, trial periods followed by referenda also helped make the process more participatory, which voters appreciated. Policy-makers also regularly provided information about the policy's effects to citizens, which also likely helped (8). In other contexts, however, information provision may be politicized, which in our view makes rigorous policy evaluation all the more important.

EMPIRICAL FRAMEWORK

Policy evaluation relies on randomized controlled trials (e.g., expanding Medicaid through a lottery for low-income households) or observational studies (e.g., of neighborhood policing). Causal analysis of public support would accompany policy evaluation whenever feasible. Randomized controlled trials often involve the use of surveys to measure outcomes and determine whether the policy had the intended effects. The cost of also measuring public support and beliefs about the policy or intervention of interest may thus be low. Econometric techniques may also allow examining policies not implemented at random. Recent econometric advances may allow exploiting policies implemented by a single jurisdiction or staggered across jurisdictions, such as abortion legalization (11). Medicaid expansion is a good example of a staggered policy that used to be unpopular among a large portion of the public, but is likely no more, despite efforts to politicize it.

The ideal framework is one where the researcher builds a survey panel, so that the

benefits while overestimating their distributional and economic impacts (1).

By systematically studying public support alongside policy evaluation, scholars would learn at the same time about the policy's effects as well as about voters' process of learning about the policy. Policy-makers are likely to care about both items. Considering early on questions of acceptability is fully compatible with "option C thinking," the idea that researchers should prioritize collecting evidence not only on how policies work in a given setting, but also on how they can be maintained and scaled up (3), which naturally includes questions of political buy-in. Our proposal complements

[1]Department of Economics, Georgia State University, Atlanta, GA, USA. [2]Department of Economics, Erasmus University Rotterdam, Netherlands. [3]Department of Economics, University of Chicago, Chicago, IL, USA. Email: scarattini@gsu.edu

same individuals in both the treatment and control groups are surveyed at least twice, before and after the treatment group receives the treatment, which in our context is the implementation of a new policy. This longitudinal approach is different from opinion polls, which rarely use control groups and generally interview different respondents across waves. As for standard policy evaluation, absent randomization, the researcher would need to identify a proper control group. The survey would ideally be representative of the underlying population of interest. Survey data could also help policy evaluation, complementing administrative data in examining behavioral change.

A trade-off between policy evaluation and causal analysis of public support may arise if the two approaches require different control groups, increasing survey costs. Researchers performing policy evaluation need to account for potential spillovers, which could "contaminate" the control group (think of a vaccine that protects also individuals who do not take it). In the case of the causal analysis of public support, this issue could be more severe. It is plausible to assume that if some cities experimented with a new policy and the policy was successful, voters in other cities could learn about it as well. In this case, the effect of experience on public support will be measured in a conservative way. If one is very concerned about spillovers, a potential solution consists in using an already-treated control group, where voters are less likely to learn something new from policy implementation in the treatment group. However, it is still possible that as a policy expands, respondents in frontrunner jurisdictions may see it more favorably, in particular if the policy tackles a global public good, such as climate change mitigation. For instance, survey respondents in British Columbia saw their carbon tax even more favorably after the implementation of a federal carbon tax in 2019 covering all Canadian provinces without their own carbon pricing scheme (*12*).

Another issue that researchers need to consider is potential experimenter demand effects, which imply that survey respondents may provide answers that they believe are going to please the researchers. Research in political science shows that experimenter demand effects in survey experiments are likely to be of limited size (*13*). However, when a survey respondent is interviewed about the same policy twice, it may happen that the second time, the respondent is more likely to provide a positive appraisal, in particular if exposed to the policy (i.e., in the treatment group) and if feeling that the researcher may expect them to like the policy more when asking again. One solution here consists in adding new respondents for the second wave of the survey, also representative of the underlying population. If first-time respondents provide responses similar to those of the second-time respondents, then experimenter demand effects are likely not an issue.

More than two survey waves may be useful, too, to measure medium- to long-run effects and pin down the exact timing of belief revisions across voters. In the aforementioned studies, belief revision is observed within 6 to 12 months from policy implementation but could have occurred earlier. Policy-makers up for reelection may be especially interested in belief revision's

> "...scholars would learn... about the policy's effects as well as about voter's process of learning about the policy."

timing. The above-mentioned procedure would be applied to multiple survey waves. Adding new respondents for each new wave would not only help to address potential experimenter demand effects, but would also maintain the sample size constant over time. Indeed, it is common for a portion of the survey respondents in the baseline survey not to respond when invited to the follow-up survey. Such attrition reduces a panel's sample size.

Although replenishing the sample is important, it is also equally important to check, on the basis of socioeconomic characteristics and other observable dimensions, that respondents attrite at random, implying no differences in characteristics between one-time and multiple-time respondents and no differences in attrition between the treatment and the control group. In the medium run, data on voting behavior may also complement survey data on public support, especially when the interest is in particularly controversial policies.

IMPLICATIONS FOR POLICY AND RESEARCH

The main implication for research consists, as discussed, in systematically combining policy evaluation with the causal analysis of public support, in particular for policies and interventions that are known to be (ex ante) unpopular. Such development in research should be encouraged by funders. Many funders prioritize causal inference and already support large randomized controlled trials, in developed and developing countries alike, sometimes testing new policies that may not enjoy widespread support, such as making teachers' salaries dependent on their attendance (*14*). In the future, they could also encourage researchers to causally evaluate public support for the policy or intervention of interest. If policy evaluation can be realized just with administrative data, funders could make available extra resources for panel surveys, including the extension of existing panels.

Funding studies in different contexts may also allow tackling questions of generalizability, which are common to both policy evaluation and the causal analysis of public support. Generalizability is especially important when evidence from front-runner jurisdictions is used to inform other jurisdictions.

Policy-makers could also facilitate the researcher's job by announcing early on plans for implementing policies, sticking to schedule, and issuing requests for proposals to examine a forthcoming policy. Policy evaluation may also help voters learn about a policy's effect. For policies that tend to have very salient effects, such as congestion charges or pricing garbage by the bag, voters' learning may happen seamlessly. For other policies such as carbon taxes, for voters to learn about their effects, it may be necessary to first have a policy evaluation exercise and its results being conveyed to voters. Hence, it may be in policy-makers' interest in front-runner jurisdictions to collaborate actively with researchers and make data available to facilitate policy evaluation. Rigorous policy evaluation may also reduce the chance of competing narratives on a policy's effect. We support calls for more experimentation and staggering whenever practical (*15*), especially as countries are investing big in new social, infrastructural, and environmental programs. ∎

REFERENCES AND NOTES

1. S. Carattini, M. Carvalho, S. Fankhauser, *Wiley Interdiscip. Rev. Clim. Change* **9**, e531 (2018).
2. W. S. Vickrey, *Am. Econ. Rev.* **53**, 452 (1963).
3. J. A. List, *Nature* **626**, 491 (2024).
4. I. Haaland, C. Roth, J. Wohlfart, *J. Econ. Lit.* **61**, 3 (2023).
5. T. L. Cherry, S. Kallbekken, S. Kroll, *Resour. Energy Econ.* **38**, 84 (2014).
6. E. Dal Bó, P. Dal Bó, E. Eyster, *Rev. Econ. Stud.* **85**, 964 (2018).
7. G. Schuitema, L. Steg, S. Forward, *Transp. Res. Part A Policy Pract.* **44**, 99 (2010).
8. D. Andersson, J. Nässén, *J. Transp. Geogr.* **52**, 82 (2016).
9. B. Murray, N. Rivers, *Energy Policy* **86**, 674 (2015).
10. S. Carattini, A. Baranzini, R. Lalive, *Ecol. Econ.* **148**, 131 (2018).
11. S. Cunningham, *Causal Inference: The Mixtape* (Yale Univ. Press, 2021).
12. M. Mildenberger, E. Lachapelle, K. Harrison, I. Stadelmann-Steffen, *Nat. Clim. Chang.* **12**, 141 (2022).
13. J. Mummolo, E. Peterson, *Am. Polit. Sci. Rev.* **113**, 517 (2019).
14. E. Duflo, R. Hanna, S. Ryan, *Am. Econ. Rev.* **102**, 1241 (2012).
15. P. J. Ferraro *et al.*, *Science* **381**, 735 (2023).