RESOURCES *for the* FUTURE

Learning How to Build Back Better through Clean Energy Policy Evaluation

Joseph E. Aldy

Working Paper 22-15
August 2022

About the Author/Authors

Joseph E. Aldy is a university fellow at Resources for the Future and a Professor of the Practice of Public Policy at Harvard's Kennedy School. His research focuses on climate change policy, energy policy, and mortality risk valuation. Aldy also currently serves as the faculty chair of the Regulatory Policy Program at the Harvard Kennedy School. In 2009–2010, he served as the special assistant to the president for energy and the environment, reporting through both the White House National Economic Council and the Office of Energy and Climate Change.

Acknowledgements

Fantastic research assistance was provided by Michael Chen, Emily Fry, Charles Hua, Michelle Li, Connor McRobert, Em Murdock, Ken Norris, Sidd Shrikanth, and Dan Stuart. I have benefitted from excellent feedback from Danielle Arostegui, Julie Gohlke, Wes Look, Kevin Rennert, Morgan Rote, Beia Spiller, and Natasha Vidangos, seminar attendees at Florida State, MIT, Resources for the Future, University of Houston, and the ASSA annual conference. This research has been supported by Environmental Defense Fund, Harvard University Center for the Environment, HKS Mossavar-Rahmani Center for Business and Government, and HKS Center for Public Leadership.

About RFF

Resources for the Future (RFF) is an independent, nonprofit research institution in Washington, DC. Its mission is to improve environmental, energy, and natural resource decisions through impartial economic research and policy engagement. RFF is committed to being the most widely trusted source of research insights and policy solutions leading to a healthy environment and a thriving economy.

Working papers are research materials circulated by their authors for purposes of information and discussion. They have not necessarily undergone formal peer review. The views expressed here are those of the individual authors and may differ from those of other RFF experts, its officers, or its directors.

Sharing Our Work

Abstract

The Infrastructure Investment and Jobs Act, the CHIPS and Science Act, and the Inflation Reduction Act authorized and appropriated unprecedented spending and tax expenditures to decarbonize the American economy. In the spirit of "build back better," this paper examines how integrating evaluation in the design and implementation of these new clean energy policies can facilitate the learning necessary for policymakers to make policy better over time. It draws lessons from two case studies: (1) on institutionalizing evaluation based on the experience with regulatory review, and (2) on conducting evaluation based on the research literature assessing the 2009 Recovery Act's clean energy programs. The paper identifies in recent legislation the programs and their characteristics amenable to various evaluation methodologies. The paper closes with recommendations for a clean energy program evaluation framework that would enable implementation of climate-oriented learning agendas under the Evidence-Based Policymaking Act.

Contents

1. Introduction **1**

2. Institutionalizing Program Evaluation: Lessons from Regulatory Review **3**

2.1. Demonstrating the Compelling Need for Policy Action 3

2.2. Standardizing Evaluation Methods and Process 4

2.3. Promoting a Culture for Retrospective Analysis and Iterative Policymaking 5

3. Conducting Program Evaluations: Lessons from Academic Research of the American Recovery and Reinvestment Act of 2009 **6**

3.1. Learning through Randomization: The Weatherization Assistance Program 7

3.2. Comparing Winners and Losers: Small Business Innovation Research Grants 9

3.3. Exploiting State Variation: State Energy-Efficient Appliance Rebate Program 11

3.4. Exploiting Formula Allocations: Employment Impacts of Clean Energy Programs 12

4. Planning for Clean Energy Program Evaluations **14**

4.1. Develop Cross-cutting and Agency-specific Guidance for Performance Evaluations 14

4.2. Identify Priority Outcomes to Evaluate 14

4.3. Identify Policies and Programs with Significant Learning Potential 15

4.4. Develop Evaluation Plans and Data Protocols 15

4.5. Ensure Evaluation Plan Transparency 16

4.6. Promote a Performance Evaluation Culture 17

5. Conclusions and Policy Implications **18**

6. References **19**

1. Introduction

In 2021, the U.S. Government pledged to reduce its greenhouse gas emissions 50-52 percent below their 2005 levels by 2030 and to achieve economy-wide net-zero emissions by 2050. To make progress on these emission goals, the Biden Administration and Congress have advanced an ambitious program to "build back better" through the Infrastructure Investment and Jobs Act,[1] the CHIPS and Science Act,[2] and the Inflation Reduction Act.[3] These laws build on decades of clean energy policy at the federal, state, and local levels, including: tax credits, accelerated depreciation, tax exemptions, rebates, grants, loans, loan guarantees, and regulatory and information disclosure requirements. In the spirit of build back better, integrating program evaluation in the design and implementation of new clean energy policies can facilitate the learning necessary for policymakers to make policy better over time: increasing the likelihood of achieving climate goals and reducing the costs of doing so.

Three key characteristics of the climate challenge illustrate the significant value in evaluating clean energy policy performance. First, transforming the modern energy economy to combat climate change will require unprecedented depth and breadth of policy action. Past policy experiences likely provide incomplete insights for how to design ambitious decarbonization policies. A continuous learning process will be needed as we deploy new technologies and policy strategies. Second, many technological, environmental, social, and economic uncertainties characterizing clean energy will be resolved by policy practice. Some policies will turn out more effective than expected, while others less effective than expected. Policy experimentation reducing uncertainty will provide the foundation for making policy better over time. Third, the policy response to climate change will continue to occur through a series of bills and regulations over time: annual appropriations; tax extender packages; agriculture, energy, and transportation bills; reconciliation bills; other legislation; regulatory standards, and more. Iterative policy processes create opportunities for using lessons to inform and improve future policy design.

Understanding the causal impacts of policy—e.g., how did a clean energy policy directly change emissions, energy investment, employment, public health, etc.—is critical for improving policy design and implementation over time. As the Commission on Evidence-Based Policymaking (2017) noted, "[p]olicymakers must have good information on which to base their decisions about improving the viability and effectiveness of government programs and policies. Today, too little evidence is produced to meet this need" (p. 1). Despite the dearth of adequate evidence, the Commission emphasized a constructive path forward: "[m]odern technology and statistical methods, combined with transparency and a strong legal framework, create the opportunity to use data for evidence building in ways that were not possible in the past" (p. 1). The Foundations for Evidence-Based Policymaking Act of 2018

1 P.L. 117-58.

2 P.L. 117-167.

3 P.L. 117-169.

reflects many of this bipartisan commission's recommendations.[4] This legislation and associated implementation institutionalizes program evaluation as a critical element of learning agendas throughout the federal government (Office of Management and Budget [OMB] 2019, 2020, 2021a). These learning agendas "identify, prioritize, and establish strategies to develop evidence to answer important short- and long-term strategic questions (i.e., questions about how the agency meets its mission(s), including about how programs, policies, and regulations function both individually and in combination)" (OMB 2019, p. 14). OMB (2021a) guidance tasks agencies to develop annual evaluation plans in their learning agendas.

Evaluating programs will provide "confidence that government investments are being used wisely to benefit the American people" (Yagan 2021). A clean energy policy performance evaluation framework would: (1) improve policy design and delivery; (2) identify new policy needs and inform policy updating; (3) enable more effective cross-agency coordination; and (4) demonstrate policy effectiveness that could enhance public and stakeholder support.

Realizing the potential for learning to inform and improve clean energy policy over time requires careful planning for program evaluation. Developing an evaluation plan at the program design stage can enhance the value of the evaluation. Such planning can ensure the collection of the necessary data for the evaluation. Committing publicly to an evaluation strategy at the design stage and publishing the results of the evaluation builds public trust in the analysis. Moreover, public commitment to future evaluation increases the likelihood that it will occur by raising the political costs of failing to do so. Institutionalizing the application of evaluation results in policy updating and strategy development will encourage a culture supporting evaluation.

This paper examines how the evaluation of clean energy and climate-oriented investment programs can spur iterative improvements in policy over time. The next section provides a case study on institutionalizing evaluations based on the experience with regulatory review. The third section presents a case study on conducting evaluations based on academic research of the American Recovery and Reinvestment Act of 2009 clean energy programs. Synthesizing the lessons from these case studies, the fourth section describes how to plan for and use the information learned from clean energy program evaluations. The final section concludes with a discussion of policy implications.

4 P.L. 115-435.

2. Institutionalizing Program Evaluation: Lessons from Regulatory Review

Since 1981, Republican and Democratic Administrations have required regulatory agencies to estimate the prospective benefits and costs of their major regulatory proposals as a part of the regulatory review process.[5] Environmental and energy regulations represent a disproportionate share of federal regulatory proposals. Over 2007-2016, the Environmental Protection Agency (EPA), Department of Energy, and Department of Transportation (in rules jointly-issued with EPA) issued more than half of all major federal regulations (OMB 2018). These environmental and energy rules represent more than 85 percent of the prospective benefits and 75 percent of the prospective costs of major Federal regulations (Aldy 2020b). The experience with regulatory review holds three major lessons for institutionalizing clean energy program evaluation.

2.1. Demonstrating the Compelling Need for Policy Action

Policymakers can communicate more effectively why a policy action is in our nation's interest by marshalling evidence of the impacts of that policy action. For example, the current regulatory review process requires federal agencies to demonstrate that their regulations address a "compelling need, such as material failures of private markets to protect or improve the health and safety of the public, the environment, or the well-being of the American people" (E.O. 12866, §1(a)). In a rulemaking, a regulatory agency identifies the market failure, highlights how the proposed regulatory action addresses the market failure and why it is preferred to alternative approaches, and shows how the benefits justify the costs. The "compelling need" standard that motivates regulatory actions would reasonably apply to any public policy, including spending and tax expenditures, that promotes clean energy investment to combat climate change. Spending and tax policy that deliver on the same objective as a regulatory action merit a comparable approach to evaluation.

Virtually all clean energy spending effectively subsidizes investment in equipment and capital that could be mandated under regulatory standards to address climate change-related market failures. For example, furnaces have been subject to minimum energy efficiency standards,[6] qualified for energy-efficient appliance rebates (Houde and Aldy

5 See: E.O. 12291, 46 Federal Register 13193, February 17, 1981; and E.O. 12866 , 58 Federal Register 51735, October 4, 1993.

6 Refer to "Energy Conservation Program for Consumer Products: Energy Conservation Standards for Residential Furnaces and Boilers," 72 Federal Register 65136, November 19, 2007.

2017), and been eligible for tax credits.[7] Wind power has been eligible for production tax credits, §1603 grants, and §1705 loan guarantees (Aldy 2013), and played a key role in determining emission standards under EPA's Clean Power Plan (Fowlie et al. 2014).

Just as analysis can inform the selection and design of preferred regulatory options, evaluations of spending and tax programs can enhance policymaker understanding of the most effective instruments for delivering on clean energy objectives. Producing such analyses take time and resources; thus, the regulatory review requirements apply only to the largest regulatory actions—those with at least $100 million in annual economic impacts—where the value of information generated is likely to be greatest. The $100 million impact threshold that triggers a full-blown analysis of regulatory impacts is modest relative to the size of major clean energy tax and spending programs in recent laws (e.g., the Infrastructure Investment and Jobs Act and the Inflation Reduction Act). These regulatory analyses matter in the regulatory development process: they inform changes to the rule after the proposal stage, and they are required to be submitted to Congress with all major final rules under the Congressional Review Act.

2.2. Standardizing Evaluation Methods and Process

The evaluation of clean energy programs can draw from existing guidance in the regulatory space. They could also draw from program evaluation procedures applied to non-climate policies in other parts of the federal government, such as the Departments of Health and Human Services and Labor. The development of standard procedures for evaluating clean energy spending programs could reduce the time and resource requirements for planning and executing program evaluations. Such standardized procedures and guidance could fall under a department's learning agenda and plan development under the Evidence-Based Policymaking Act.

For example, OMB (2003) issues guidance to regulatory agencies on the conduct of regulatory impact analyses. The guidance addresses the economic principles and some common economic assumptions that should inform agency estimation of benefits and costs. The guidance emphasizes both the expected rigor of analysis—and the importance of relying on peer-reviewed literature—as well as the communication of the results of the analysis to enable a clear understanding by policymakers, stakeholders, and the public. Such regulatory impact analyses often go beyond simply tallying and comparing benefits and costs; they also present estimated employment and competitiveness impacts, ancillary benefits beyond the target of the rule, as well as the distribution and uncertainty characterizing the impacts of the regulatory action (Aldy et al. 2021, Robinson et al. 2016).

Several regulatory agencies have developed their own guidance for the conduct of prospective regulatory impact analyses, such as EPA (2014) and Department of

7 P.L. 111-5, section 1121.

Health and Human Services (2016). The Department of Transportation (2021) issues regular updates of its approach for valuing reductions in mortality risk through its regulatory authorities. The Biden Administration relaunched the interagency working group on the social cost of greenhouse gases, which provides estimates of the social cost of carbon, methane, and nitrous oxide that can monetize the benefits of reducing greenhouse gas emissions through regulation and other Federal actions.[8] To improve understanding of the environmental justice implications of federal investments, OMB (2021b) issued guidance for how to calculate and report the benefits of such actions under the Justice40 Initiative. These guidance documents typically have undergone peer review, such as through the EPA Science Advisory Board, the National Academies, and other processes.

2.3. Promoting a Culture for Retrospective Analysis and Iterative Policymaking

The sunset provisions for clean energy spending and tax expenditures through the Infrastructure Investment and Jobs Act and the Inflation Reduction Act create windows of opportunities for how looking back at program performance can inform subsequent policy actions. Likewise, the iterative approach to regulations creates natural opportunities for ex post evaluation of regulatory performance. A number of regulatory authorities operate through an updating cycle, such as EPA air quality standards,[9] Department of Energy appliance efficiency standards,[10] and Department of Transportation fuel economy standards.[11] Looking back at regulatory performance provides an opportunity to learn about the efficacy of rule design and compliance strategies by regulated entities, and significantly enhances knowledge of regulatory impacts relative to the prospective analysis developed at the rule-writing stage (Greenstone 2009, Sunstein 2011, Aldy 2014a).

Regulatory agencies' practice with respect to retrospective review of existing regulations—which would be analogous to a clean energy program evaluation framework—has yielded a mixed record (Harrington 2006, Coglianese 2013, Aldy 2014a, Bull 2015, Cropper et al. 2017). Every administration dating back to the Carter Administration has called on regulatory agencies to review their existing rules, but the failure to meaningfully institutionalize retrospective review, build a culture of such review within agencies, and appropriate monies to ensure the resources are available to conduct such reviews, have undermined the effectiveness of such White House directives. Agencies have received guidance on how to plan for ex post evaluations of regulations during the rule-making stage, but few have moved forward with such strategies (ACUS 2014, Aldy 2014a, Cropper et al. 2018).

8 E.O. 13990, 86 Federal Register 7037, January 25, 2021.

9 42 USC 7409(d).

10 42 USC 6313(a)(6(C).

11 49 USC 32902(k)(3).

Promoting a culture for retrospective analysis starts with institutionalizing its use by political leaders and the policy process. If there is neither an obvious audience for the analysis nor a process for using the outputs of the analysis for improving policy, then agencies will consider such evaluations of policies in practice a low priority. During the Obama Administration's retrospective review effort, agencies posted online the list of rules under review and the results of those reviews. Over time, however, these periodic updates by regulatory agencies received less attention from the White House, stakeholders, and the media (Aldy 2014a).

3. Conducting Program Evaluations: Lessons from Academic Research of the American Recovery and Reinvestment Act of 2009[12]

The challenge in learning about policy impacts lies in identifying the appropriate data and implementing the rigorous evaluation tools to produce a robust understanding of the impact of clean energy programs. A program evaluation is much more than simply reporting the number of participating firms or households in a program, or taking such a count and multiplying it by an engineering-based outcome, such as expected energy savings. Empirical social scientists have developed an array of evaluation tools—field experiments that implement randomized control trials as well as quasi-experimental methods that attempt to replicate the fundamental characteristics of a randomized control trial (e.g., Angrist and Pischke 2008, Lee and Lemieux 2010, DiNardo and Lee 2011, Imbens and Rubin 2015)—to estimate the various outcomes caused by a program or policy intervention.

Estimating the causal impact of a clean energy program requires information about both those who participate in the program and those who do not. Simply collecting data from those receiving grants or claiming tax credits would be insufficient; rigorous analysis also depends on data about those households and businesses that are similar to the subsidy recipients but are not recipients. These non-participant data provide the basis for the counterfactual—what would have happened in the absence of the policy—that enables analysis of program performance. In effect, data on program participants represents information on a "treatment" group and data on non-participants represents the information on a "control" group, just as in a randomized experiment to evaluate the impacts of a drug or vaccine.

12 For general assessments of the Recovery Act's clean energy package, refer to Aldy (2013), Carley (2016), and Barbier (2020).

The ambitious spending and policy experimentation under the American Recovery and Reinvestment Act of 2009 has been subject to extensive program evaluations in the academic literature. The Recovery Act provided about $100 billion in clean energy spending and tax expenditures to promote deployment of low-carbon technologies and spur economic activity (Aldy 2013, CEA 2016). The energy landscape has change dramatically since the Recovery Act was signed into law in February 2009: utility-scale solar power generation is more than 100 times greater and wind power generation is nearly seven times greater today than in 2008 (EIA n.d.). Policymakers could draw from this past experience in evaluating Recovery Act programs to apply program evaluation methods to new clean energy policies going forward.

This section presents illustrations of methods for conducting program evaluations that credibly estimate the causal impacts of clean energy programs. I show how each of these methods can be applied using studies of four clean energy programs supported by the Recovery Act. In each case, I open by describing the potential biases that may result in misleading claims of program performance based on program participation rates and engineering assumptions. Then I describe the authors' study and application of a statistical method that can account for and minimize these biases. For each case study, I note how the study's method could inform future program evaluations for specific clean energy programs in the Infrastructure Investment and Jobs Act and the Inflation Reduction Act.

3.1. Learning through Randomization: The Weatherization Assistance Program

The 2009 Recovery Act provided nearly $5 billion of funding for Weatherization Assistance Programs (WAP) implemented at the state and local levels. These weatherization programs finance energy-efficiency and conservation improvements in the residential dwellings of households with income below a specified threshold.

3.1.1. Potential Biases

The Department of Energy has typically estimated the reduced energy demand and associated energy bill savings of weatherization through engineering-based evaluations (e.g., Oak Ridge National Laboratory 2015). Engineering-based analyses suffer from three potential shortcomings. First, the weatherization investment in practice may yield different energy savings because of simplifying assumptions in the engineering model or variations in the quality of the contractors undertaking the work. Second, individuals opting to participate in a weatherization program may be fundamentally different—perhaps they are more energy or environmentally conscious—from the general population, and their behavior may not be representative. Finally, weatherization lowers the cost of an energy service—such as heating a home to a given temperature. Residents of a weatherized home may adjust the thermostat, or buy more energy-consuming appliances, and this so-called "rebound effect" would offset some of the energy savings.

3.1.2. An Evaluation Strategy to Address the Biases

In policy debates, there has occasionally been a tension between advocates of program evaluation—who argue for implementing a public program through a randomized control trial to enable rigorous assessment—and agency staff or politicians who claim that the program should be available to everyone who is eligible. Fowlie et al. (2018) developed a clever way of resolving this tension. Working with a local weatherization program in Michigan, they developed a randomized encouragement program—they did not alter who was eligible for this means-tested program, but they randomized who received information and technical assistance for applying for weatherization aid. This randomization satisfied political constraints, and also allowed the researchers to ensure that their results were not confounded by, for example, self-selection into the program by those more likely to be energy-conscious. They also acquired the pre- and post-weatherization building-specific energy consumption data to enable them to compare estimated energy savings to the engineering model results.

They found that providing information and assistance to a randomly-selected set of households increased these households' participation in the program, but they also found much smaller energy savings than estimated in engineering models. To avoid the risk that a household predisposed to want to save energy would participate in WAP—and potentially bias the estimated impacts of weatherization—the researchers employed this randomized encouragement in their statistical model as a way to ensure that the comparison of outcomes, such as energy consumption, between WAP participants and non-WAP participants was based on who randomly received information about WAP programs. While they estimated WAP caused energy consumption to decline 10-20 percent for participating households, these energy savings were about two-thirds less than engineering estimates.

3.1.3. Opportunities to Apply this Strategy to Future Clean Energy Programs

Randomized encouragement can apply to clean energy programs without altering program eligibility rules. Like a standard experimental study, eligible individuals can be assigned randomly to treatment (receiving additional information and assistance in applying for weatherization) or control groups. Given scarce resources in publicizing clean energy programs, program managers may have an opportunity to use these resources in a strategic manner to enable policy learning in the future.

Learning that the weatherization program in practice fell nearly two-thirds short of the engineering model energy savings motivated subsequent research to examine this finding. In an evaluation of Illinois weatherization assistance program projects, Christensen et al. (2021) employ recent advances in machine learning techniques to decompose the wedge between engineering and empirical program evaluation estimates of weatherization. They find that a modest amount (6 percent) of this gap reflects the behavioral response of the homeowners (the rebound effect), while overestimated savings in the engineering models, especially for insulation, and significant heterogeneity in contractor quality each represent more than 40 percent of the gap.

Such randomized encouragement and randomized control trials could enable evaluation of an array of Inflation Reduction Act programs. For example, the Rural Energy for America Program could randomly provide farmers and ranchers information on and technical assistance for applying for grants and loan guarantees for renewable power and energy efficiency investments. The Greenhouse Gas Reduction Fund could strategically provide technical assistance, in a randomized manner, as a part of its grant-making for clean energy technologies. State agencies will receive an array of block grants from EPA and the Department of Energy, and they could deploy some monies for randomizing the targeting of information about §25C energy efficiency home improvement tax credits, §25D residential clean energy tax credits, and §30D clean vehicle tax credits. In addition, such experiments could allow for evaluations of how technology deployment allows for subsequent policy reforms. For example, Fowlie et al. (2021) employed an experiment that took advantage of smart meter deployment under the Recovery Act to randomly opt households into time-varying electricity pricing in order to evaluate the efficiency benefits of pricing reforms.

3.2. Comparing Winners and Losers: Small Business Innovation Research Grants

In the 2009 Recovery Act, the federal government provided additional resources for, inter alia, energy-related innovation through the Small Business Innovation Research (SBIR) grant program. The SBIR program has been a long-running federal program that requires a specified percentage of a department's research grant budget to target small businesses, which, in the context of the Department of Energy, can include many start-ups attempting to commercialize novel, low-carbon technologies.

3.2.1. Potential Biases

The challenge in assessing the efficacy of spending on innovation is that agencies typically track the outcomes of only the recipients of the spending. If the recipient files for and receives a patent, then the agency may associate that patent as a result of the funding. The advance in knowledge represented in the patent, however, may have occurred even in the absence of federal funding. This could reflect a very high-quality recipient who was already on the cusp of the innovation before receiving the federal funding. Or the public funding of innovation may crowd out private funding that would have occurred in the absence of the public program. The bottom line is that observing only the outcomes for innovation grant recipients is insufficient to illustrate the impact of an innovation program.

3.2.2. An Evaluation Strategy to Address the Biases

Using data on the winners and losers of hundreds of energy SBIR grant competitions over 1983-2013, Howell (2017) evaluated the impacts of providing grants to small businesses developing novel energy technologies on patenting activity and on the likelihood of receiving additional venture capital. In conducting these grant competitions, a program official at the Department of Energy ranks all projects within a competition, and then, after the ranking has been submitted to the SBIR office, another official determines the cutoff in the ranking—those above the cutoff receive grants—as a function of broader budget constraints. Howell acquired the full rankings for these energy grant competitions and compared outcomes for the lowest-ranked grant recipient to the highest-ranked non-recipient. Since the determinations of the cutoffs were not related to the rankings, the assigning of the grant to one (the lowest-ranked recipient) but not the other (the highest-ranked non-recipient) could be plausibly considered random. By focusing on pairs of small businesses—in each pair, one just above the cutoff and one just below—Howell could estimate how grant receipt influenced small businesses' patenting and financing outcomes relative to non-receipt by otherwise similar small businesses. She finds that the SBIR phase 1 grant awards increased cite-weighted patents by 30 percent and nearly doubled the probability that a small business would receive future venture capital funding.

3.2.3. Opportunities to Apply this Strategy to Future Clean Energy Programs

This empirical strategy would map well to many competitive grant programs throughout the government. In the context of energy innovation, such an approach of comparing grant winners to grant losers has also been used to show how ARPA-E funding recipients filed patents at twice the rate of comparable firms rejected by ARPA-E (Goldstein et al. 2020). The §48C clean energy manufacturing tax credit under the 2009 Recovery was a competitive tax credit capped at $2.3 billion of tax expenditures implemented through a ranking process conducted by the Departments of Energy and Treasury. With $10 billion allocated to the §48C tax credit in the Inflation Reduction Act, a series of annual solicitations of projects (e.g., 10 $1 billion competitions) could be evaluated through a comparison of lowest-ranked recipients and highest-ranked non-recipients. Moreover, the Infrastructure Investment and Jobs Act's $3.5 billion of support for regional direct air capture hubs and the Inflation Reduction Act's $5.8 billion for the Advanced Industrial Facilities Deployment Program could each be implemented through a competitive selection process amenable to such an analysis (also see Greenstone et al. 2010 for a similar type of analysis of the impacts of siting large manufacturing facilities). In general, any clean energy program in which a competition determines a winner and identifies the next-best alternative could be evaluated within such a framework. Finally, such an approach could also examine the impacts of programs that target specific communities, such as low-income and disadvantaged communities under the Greenhouse Gas Reduction Fund, by comparing outcomes in communities on each side of the metric that determines community eligibility for the program.

3.3. Exploiting State Variation: State Energy-Efficient Appliance Rebate Program

The Recovery Act of 2009 relied on partnerships with state agencies to implement many clean energy programs. As states design and implement their energy programs, there may be a number of variations that can serve as the basis for program evaluation, such as in the performance of the State Energy Efficient Appliance Rebate Program. The 2009 Recovery Act represented the first appropriation of monies to implement this rebate program.

3.3.1. Potential Biases

A number of clean energy spending programs aim to accelerate investment that is already occurring in the economy. This presents a challenge in program design: how to target spending to those who would not have undertaken the investment anyway in the absence of the program. Agencies' reporting on Recovery Act programs often gave the impression that all energy investment associated with a program was marginal, i.e., occurred only because of the existence of the programs. The Department of Energy's assessment of the state rebate program for EnergyStar-rated appliances suggests that all rebate claimants would have bought a non-EnergyStar appliance or continued using a less-efficient appliance in the absence of the program (Department of Energy 2015). With EnergyStar appliance market shares in excess of 50 percent for refrigerators, dishwashers, and clothes washers before the start of these rebate programs, there were already many households purchasing rebate-eligible products.

3.3.2. An Evaluation Strategy to Address the Biases

Under the Recovery Act, state governments developed new programs to administer rebates for EnergyStar-rated appliances, subject to Department of Energy review and approval. In practice, states implemented programs that varied in terms of program timing, types of eligible appliances, and rebate amounts. Houde and Aldy (2017) exploited state variation to estimate the market and energy impacts of this program. In effect, a state with a rebate program that is on for a given appliance could be compared to another state program that is off for that appliance. They evaluate the rebate programs for refrigerators, dishwashers, and clothes washers using data secured from a national retailer to estimate how state rebate programs caused changes in appliance markets and expected consumer energy consumption through appliance purchases. Houde and Aldy find that as many as 90 percent of the EnergyStar appliances purchased under this program would have happened anyway. This reflected households already planning to buy an eligible appliance as well as a short-term shifting in the timing of an eligible appliance purchase that households would have bought anyway. With large, inframarginal claims of rebates, they estimate that the refrigerator program yielded 1-2 kilowatt-hours of electricity savings annually per rebate.

3.3.3. Opportunities to Apply this Strategy to Future Clean Energy Programs

Houde and Aldy also show how counterfactual energy efficiency subsidy policies—such as providing rebates based on energy-efficiency thresholds more ambitious than the minimum necessary under EnergyStar—could significantly improve the energy savings and cost-effectiveness of the program. This shows how such evaluations can inform our understanding of past program performance and illustrate ways of improving performance through program revision.

This empirical evaluation strategy represents a common approach to estimating the causal impacts of energy and environmental regulations and subsidies (e.g., Aldy et al 2022 review Clean Air Act regulatory performance evaluation studies and Gillingham et al 2018 review evaluations of energy efficiency policies and programs) and many health, labor, and public programs (e.g., Angrist and Pischke 2008, DiNardo and Lee 2011). Continued reliance on federal-state partnerships implementing energy and environmental programs will enable evaluators to exploit variation in state programs (the "laboratories of democracy") to estimate the impacts of those programs. For example, an evaluation of diesel school bus retrofits that exploited the variation in program participation across school districts in Georgia over time found that it improved children's respiratory health and test scores (Austin et al. 2019). The Inflation Reduction Act appropriated $1 billion for additional school bus and other heavy duty vehicle diesel retrofits. State grants through the Greenhouse Gas Reduction Fund could also be evaluated through this strategy. Finally, more general spatial and temporal variation in the implementation of clean energy programs can serve as the basis for this approach, such as in the analysis of the cash for clunkers program by Mian and Sufi (2012).

3.4. Exploiting Formula Allocations: Employment Impacts of Clean Energy Programs

As a part of a major economic stimulus bill, the clean energy package included the dual objectives of advancing the deployment of clean energy technologies and creating new jobs (Aldy 2013).

3.4.1. Potential Biases

Estimating the impacts of spending programs on employment can be difficult when accounting for variation in underlying economic conditions. For example, targeting spending to communities with higher levels of unemployment may be effective at spurring job creation, but the correlation—high spending in areas with high unemployment—may mask this impact, and could even be interpreted incorrectly that spending causes unemployment. In addition, given that firms hire most new employees independent of public spending, simply counting new hires at firms that claim tax credits or grants from clean energy programs may overestimate the impact of the policy.

3.4.2. An Evaluation Strategy to Address the Biases

To address such a challenge, a number of evaluations of the Recovery Act exploited the formulas used to allocate spending across the states (Chodorow-Reich 2019). Popp et al. (2020) employed the formulas associated with Department of Energy programs to estimate the impact of clean energy spending on job creation. By estimating a statistical model that focuses on the variation in spending as a function of a formula set before the Great Recession, the analyst can avoid making the mistake of conflating the underlying economic conditions of an area receiving spending with the causal impact of the spending on that area's job creation. This would minimize the risk that spending in high unemployment areas could be misinterpreted to suggest that spending causes job loss.

In focusing on clean energy spending, Popp et al. find that clean energy programs increased employment by about 15 jobs per million dollars of spending, although these programs' job creation occurred more slowly than through the Recovery Act's non-energy spending. The job creation typically occurred in those communities with a greater prevalence of workers with occupational skills associated with "green jobs." Most of the job creation focused on manual labor occupations, with about half going into construction or waste management activities.

3.4.3. Opportunities to Apply this Strategy to Future Clean Energy Programs

Many existing federal programs employ formulas for allocating resources among the states and, in some cases, states use formulas for allocating spending to the local level. For example, Leduc and Wilson (2017) used the formulas that allocate highway grant funding to the states to investigate the impact on overall transportation investment under the 2009 Recovery Act. Currently authorized programs at the Departments of Energy and Transportation regularly use formulas for distributing funds, and some of the recently authorized programs under the Infrastructure Investment and Jobs Act and the Inflation Reduction Act do as well.

4. Planning for Clean Energy Program Evaluations

As agencies begin implementing clean energy spending and tax expenditure programs under the Infrastructure Investment and Jobs Act, the Inflation Reduction Act, and the CHIPS and Science Act, as well as new rulemakings that influence energy-related investment—they can build performance evaluation into policy implementation and integrate the insights from evaluations into agencies' learning agendas. This section synthesizes the lessons from the case studies to provide recommendations for a clean energy and climate learning agenda in the federal government.

4.1. Develop Cross-cutting and Agency-specific Guidance for Performance Evaluations

The executive branch can draw from the experiences with regulatory review and the implementation of the Foundations for Evidence-Based Policymaking Act to develop guidance documents in planning and conducting performance evaluations. OMB could coordinate with the White House Climate Change Task Force and the Council of Economic Advisers, akin to previous efforts on the Circular A-4 guidance on regulatory impact analysis, to develop such guidance documents. This effort could tap expertise throughout the federal government on: (1) designing and implementing clean energy and climate-related programs, policies, and rules; (2) collecting and publishing data and analysis at federal statistical agencies; and (3) undertaking program evaluation and economic analysis of public policies. White House agencies could solicit input from relevant scholars and stakeholders and make draft guidance available for public comment. To the extent feasible, OMB should synchronize such efforts with the guidance for agencies' learning agendas and use climate and clean energy as an example of how to design and operationalize the learning agenda framework. OMB could also task agencies to develop agency-specific, peer-reviewed guidance for planning and conducting evaluations.

4.2. Identify Priority Outcomes to Evaluate

Agencies should identify the priority outcomes that merit measurement and analysis, which would map to the agency's learning agenda. OMB's learning agenda guidance specifically calls on each agency to bring evidence to bear on strategic questions about its mission, and this effort could focus on an agency's strategies as they relate to clean energy and climate change. This will also contribute to a culture for evaluation, by connecting program evaluation to agency mission. Potential priority outcomes could include greenhouse gas emissions reductions, deployment of zero-carbon energy technologies, aggregate costs and costs per ton of emissions avoided, the benefits of improved air quality, and the distribution of these outcomes across socio-demographic characteristics, regions, and industries.

4.3. Identify Policies and Programs with Significant Learning Potential

Clean energy and climate-related programs vary by policy tool, implementing agency, industry covered, technology supported, etc. This variation suggests that there are substantial opportunities for learning, but also heterogeneity in the learning potential. As agencies move forward with implementing a clean energy program performance evaluation framework, they could focus initial evaluations on those programs and policies with the greatest potential for learning.

Agencies may target evaluations for those programs that are likely to operate over a long horizon, so that periodic evaluations could inform program revision and enhance program efficacy over time. Agencies may evaluate those programs to enable learning that could be used in the design and implementation of related policies and programs. Positive learning spillovers could also apply to programs and policies implemented by other agencies, and multi-agency coordination on learning could help identify and exploit these opportunities. Some programs and policies are more likely to benefit from learning simply because they have larger appropriations or tax scores. Thus, an expenditure trigger, akin to the economic impact threshold for regulatory impact analysis, could direct evaluation efforts to the largest programs.

4.4. Develop Evaluation Plans and Data Protocols

To build agency expertise and to promote a culture of clean energy program evaluation, agencies should create evaluation teams. These should draw from current in-house expertise and grow over time as the agency recruits new staff with training on and experience with rigorous program evaluation methods. Exploring opportunities for hosting staff with expertise from other agencies (e.g., statistical agencies, or those agencies with more experience in program evaluation) can enable a quicker ramping of evaluation capacity within the agency and spur learning about operationalizing performance evaluation.

The design of the evaluation plans should occur contemporaneously with the design of clean energy programs, policies, and rules. An evaluation plan team could account for and, where agency discretion exists, influence the design of the program to enable robust evaluation. Guidance could favor program design that facilitates rigorous evaluation of program performance so long as doing so does not undermine the program's objectives. As illustrated above, researchers have found many ways of implementing randomized control trials and applying quasi-experimental methods to estimate the causal impacts of these policy interventions.

An agency should develop a data-collection protocol consistent with the data needs of the evaluation plan. Historically, agencies collect information only on program participants or regulated entities, and they rarely collect any information associated with tax expenditures for evaluation purposes. This will require a significant rethinking of information collection, which suggests that a well-resourced and proactive Office of Information and Regulatory Affairs at OMB could play a key role through its oversight of the Paperwork Reduction Act. The data-collection protocol could establish ways to cleanly match agency-collected data with relevant data collected by other agencies and private sector data.

As many programs under the Inflation Reduction Act operate through the tax code, developing a data collection and management system for relevant tax data will be important to enable evaluation of these programs. The typical information reported on tax forms for claiming energy tax credits is insufficient for evaluating the performance of the tax credit. For example, a taxpayer claiming an investment tax credit for solar power simply needs to indicate the cumulative investment costs for all solar properties in a given tax year on IRS Form 3468. In order to learn the impact of the tax credit on capacity investment, generation, emissions, local public health, employment, and other outcomes of interest, an analyst would need more granular data on the properties in question—location, investment costs, and tax expenditures per property—that could be integrated with other Federal government databases. The Census Data Linkage Infrastructure could be expanded to enable clean energy program evaluations—by integrating IRS data, augmented to enable facility-level matching, with data compiled by the Census, the Energy Information Administration, the Environmental Protection Agency, the National Center for Health Statistics, and more. Given the revolution in big data and the statistical tools for analyzing large databases, establishing ways of integrating federal data with private database would further enable evaluations, especially place-based and distributional impacts central to the Justice40 initiative.

The reporting requirements for the Treasury Department's §1603 grant program, which renewable power producers could claim in lieu of the investment tax credit under the 2009 Recovery Act, could serve as a starting point for identifying additional information to be collected by the IRS. Aldy et al. (2022) illustrate how §1603 grant program reporting data facilitate an evaluation and comparison of the impacts of investment and output subsidies on renewable power generation.

In some cases, a federal clean energy program may delegate implementation to a state or local agency. This may create an opportunity for the federal government to task the recipient of the funds to develop their own evaluation plans, and the federal agency could provide a model plan for clean energy program performance evaluation. This form of experimentation may help evaluators learn how to learn—to assess and compare evaluation strategies for producing information on key outcomes of interest to the agency.

4.5. Ensure Evaluation Plan Transparency

Agencies typically issue preliminary guidance on new programs or proposed regulations for public comment. At this stage, an agency could also publish draft evaluation plans and solicit feedback from stakeholders and outside experts. This approach would also make more salient the key objectives of the program or policy, since they would be the focus of the performance evaluation. The agency could make the revised, final evaluation plans available through a government, online repository, drawing from the recent experience with pre-analysis plans for randomized control trials in economics research.[13] Making the evaluation plan public before program implementation reduces the opportunity in the future for an agency to "cook the books" in its program evaluation.

The agency should also develop a plan for how to communicate the results of its evaluations. This is consistent with OMB guidance on both how to communicate the results of regulatory impact analyses for regulations and how to use evidence to improve public policy through learning agendas. The publishing of results should highlight the key findings, identify outstanding questions, and address how the performance evaluation could improve the design and implementation of policy over time, including potential statutory changes for consideration by Congress.

4.6. Promote a Performance Evaluation Culture

Providing resources for program evaluations and integrating them into agency learning agendas can further institutionalize their use, attract the best agency staff to work on these efforts, and promote a culture of review and learning. Where the executive branch has the authority, the learning agendas can integrate the program evaluations into a formal, iterative process for revising the design and implementation of clean energy programs. In the cases where the executive branch lacks such discretion, it could work on legislative provisions with Congress that would formally integrate a periodic review and policy updating mechanism for clean energy programs and policies (e.g., Aldy 2020a).

A consistent approach to such communication—based on cross-cutting government-wide guidance and agency-specific guidance—could also enhance the value of the evaluation for policymakers, stakeholders, the media, and the public. Specifically, agencies could report through OMB to produce a common set of metrics, such as emissions reduced, dollars per ton of emissions avoided, improvement in public health, and distributional tables, to facilitate a whole-of-government assessment of progress. An annual report to Congress presenting a scorecard of clen energy progress across agencies would signal the importance of this work to future political appointees and agency staff. Integrating scorecard information into annul budget requests could also demonstrate the value in producing evaluations.

In addition to communicating the key results, agencies should explore ways of making as much of the data and their methods available online. This would create opportunities for outside experts to replicate and extend on the agency's analyses. Such extensions may identify new outcomes of interest, or modifications to program design, that could complement the agency's performance evaluation and signal additional ways of improving clean energy program implementation. Engagement with academic researchers could also enable the training of future agency staff who would design and execute evaluations. Such transparency may also enhance trust in the government's own evaluation and draw broader political support for more ambitious clean energy policies over time.

13 Refer to the American Economic Association Randomized Control Trial Registry at: **https://www.socialscienceregistry.org/site/about** and Miguel (2021). Also refer to the NIH clinical trial registry at **https://clinicaltrials.gov/**.

5. Conclusions and Policy Implications

Producing evidence on the performance of clean energy programs under the Infrastructure Investment and Jobs Act, the CHIPS and Science Act, and the Inflation Reduction Act can play a key role in the evolution of U.S. climate change policy. With the iterative nature of appropriations, tax expenditures, and regulations, there will naturally be opportunities for drawing from such evidence to improve the design and implementation of these policies. Integrating such efforts with agencies' broader learning agendas will institutionalize the cycle of act-learn-act and further institutionalize climate change in agencies' strategic plans. Drawing from evidence to update policies will make them more effective in delivering on the various objectives of energy and climate policy.

In addition to serving as the basis for improving domestic clean energy and climate change policy, such evaluations may influence how the United States engages the world in combatting climate change. Clean energy program evaluations would produce evidence on U.S. progress in implementing its nationally determined contribution under the 2015 Paris Agreement. This evidence could inform the reporting required under the transparency mechanism of the agreement. As the United States develops more rigorous evidence on the effectiveness of specific clean energy policies, we may export our most effective policies to other countries through bilateral engagement and by informing guidance at multilateral development banks, the Green Climate Fund, and the International Monetary Fund. In showing the most effective ways for cutting emissions, such exporting of evidence and policy could make it easier for recipient countries to pledge and deliver on ambitious decarbonization goals (Aldy 2014b).

As we innovate with new programs and policies to combat climate change, we need to bring the latest innovations in how we evaluate public policy to ensure that we are securing the greatest possible reduction in climate change risk possible for our efforts. Producing evidence on the performance of the nation's next steps towards decarbonizing the economy will be critical in demonstrating to the public, stakeholders, and policymakers that the nation is building back better. Showing how decarbonization will be, in part, a learning process in which we learn how to build back progressively better over time could broaden and make more durable public and political support for ambitious climate change policy.

6. References

Administrative Conference of the United States (ACUS). 2014. Recommendation 2014-5: Retrospective Review of Agency Rules. *79 Federal Register* 75114.

Aldy, J. E. 2013. A Preliminary Assessment of the American Recovery and Reinvestment Act's Clean Energy Package. *Review of Environmental Economics and Policy* 7(1): 136-155.

Aldy, J. E. 2014a. Learning from Experience: An Assessment of the Retrospective Reviews of Agency Rules and the Evidence for Improving the Design and Implementation of Regulatory Policy. Report prepared for Administrative Conference of the United States, November 18.

Aldy, J.E. 2014b. The Crucial Role of Policy Surveillance in International Climate Policy. *Climatic Change* 126(3-4): 279-292.

Aldy, J. E. 2020a. Carbon Tax Review and Updating: Institutionalizing an Act-Learn-Act Approach to U.S. Climate Policy. *Review of Environmental Economics and Policy* 14(1): 76-94.

Aldy, J. E. 2020b. Evaluating Regulatory Performance: Learning from and Institutionalizing Retrospective Analysis of EPA Regulations. *Case Western Reserve Law Review* 70(4): 971-1006.

Aldy, J.E., M. Kotchen, M. Evans, M. Fowlie, A. Levinson, and K. Palmer. 2021. Co-benefits and Regulatory Impact Analysis: Theory and Evidence from Federal Air Quality Regulations. *Environmental and Energy Policy and the Economy* 2: 117-156.

Aldy, J.E., Auffhammer, M., Cropper, M., Fraas, A., & Morgenstern, R. 2022. "Looking Back at 50 Years of the Clean Air Act. *Journal of Economic Literature* 60(1): 179-232.

Aldy, J.E., Gerarden, T.D. & Sweeney, R.L. 2022. Investment versus Output Subsidies: Implications of Alternative Incentives for Wind Energy. *Journal of the Association of Environmental and Resource Economists*, forthcoming.

Angrist, J. D., & Pischke, J. S. 2008. Mostly Harmless Econometrics. Princeton University Press.

Austin, W., Heutel, G., & Kreisman, D. 2019. School bus emissions, student health and academic performance. *Economics of Education Review* 70: 109-126.

Barbier, E. B. 2020. Greening the post-pandemic recovery in the G20. *Environmental and Resource Economics* 76(4): 685-703.

Bull, R.T. 2015. Building a Framework for Governance: Retrospective Review and Rulemaking Petitions. *Administrative Law Review* 67: 265.

Carley, S. 2016. Energy Programs of the American Recovery and Reinvestment Act of 2009. *Review of Policy Research* 33(2): 201-223.

Chodorow-Reich, G. (2019). Geographic cross-sectional fiscal spending multipliers: What have we learned? *American Economic Journal*: Economic Policy 11(2): 1-34.

Christensen, P., Francisco, P., Myers, E., & Souza, M. 2021. Decomposing the wedge between projected and realized returns in energy efficiency programs. *The Review of Economics and Statistics*, forthcoming.

Coglianese, Cary. 2013. Moving Forward with Regulatory Lookback. *Yale Journal on Regulation* 30: 57-66.

Commission on Evidence-Based Policymaking. 2017. The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policymaking. Washington, DC: Bipartisan Policy Center.

Council of Economic Advisers. 2016. A Retrospective Assessment of Clean Energy Investments in the Recovery Act. Washington, DC: Executive Office of the President.

Cropper, M., A. Fraas, and R. Morgenstern. 2017. Looking Backward to Move Regulations Forward. *Science* 355 (6332): 1375—76.

Cropper, M. L., R. D. Morgenstern, and N. Rivers. 2018. Policy Brief: Facilitating Retrospective Analysis of Environmental Regulations. *Review of Environmental Economics and Policy* 12 (2): 359—70.

Department of Energy. 2015. State Energy-Efficient Appliance Rebate Program: Volume 2—Program Results. Washington, DC: DOE.

Department of Health and Human Services. 2016. Guidelines for Regulatory Impact Analysis. Washington, DC: HHS Office of the Assistant Secretary for Planning and Evaluation.

Department of Transportation. 2021. Guidance on the Treatment of the Economic Value of a Statistical Life (VSL) in U.S. Department of Transportation Analyses—2021 Update.

DiNardo, J. & D.S. Lee. 2011. Program Evaluation and Research Designs. *Handbook of Labor Economics*, Volume 4A. Elsevier.

Energy Information Administration. n.d. Electricity Data Browser. URL: **https://www.eia.gov/electricity/data/browser/**.

Environmental Protection Agency. 2014. *Guidelines for Preparing Economic Analyses*. Washington, DC: National Center for Environmental Economics.

Fowlie, M., Goulder, L., Kotchen, M., Borenstein, S., Bushnell, J., Davis, L., ... & Wolfram, C. 2014. An economic perspective on the EPA's Clean Power Plan. Science 346(6211): 815-816.

Fowlie, M., Greenstone, M., & Wolfram, C. 2018. Do energy efficiency investments deliver? Evidence from the weatherization assistance program. *Quarterly Journal of Economics* 133(3): 1597-1644.

Fowlie, M., Wolfram, C., Baylis, P., Spurlock, C. A., Todd-Blick, A., & Cappers, P. 2021. Default effects and follow-on behaviour: Evidence from an electricity pricing program. *Review of Economic Studies* 88(6): 2886-2934.

Gillingham, K., Keyes, A., & Palmer, K. 2018. Advances in evaluating energy efficiency policies and programs. *Annual Review of Resource Economics* 10: 511-532.

Goldstein, A., Doblinger, C., Baker, E., & Anadón, L. D. 2020. Patenting and business outcomes for cleantech startups funded by the Advanced Research Projects Agency-Energy. *Nature Energy* 5(10): 803-810.

Greenstone, Michael. 2009. Toward a Culture of Persistent Regulatory Experimentation and Evaluation. In: *New Perspectives on Regulation*, David Moss and John Cisternino, eds. Cambridge, MA: The Tobin Project.

Greenstone, M., Hornbeck, R., & Moretti, E. 2010. Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy* 118(3): 536-598.

Harrington, Winston. 2006. Grading Estimates of the Benefits and Costs of Federal Regulation: A Review of Reviews. Resources for the Future Discussion Paper 06-39.

Houde, S., & Aldy, J.E. 2017. Consumers' response to state energy efficient appliance rebate programs. *American Economic Journal: Economic Policy* 9(4): 227-55.

Howell, S. T. 2017. Financing innovation: Evidence from R&D grants. *American Economic Review* 107(4): 1136-64.

Imbens, G. W., & Rubin, D. B. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Leduc, S., & Wilson, D. 2017. Are State governments roadblocks to federal stimulus? Evidence on the Flypaper Effect of Highway Grants in the 2009 Recovery Act. *American Economic Journal: Economic Policy* 9(2): 253-92.

Lee, D. S., & Lemieux, T. 2010. Regression discontinuity designs in economics. *Journal of Economic Literature* 48(2): 281-355.

Mian, A., & Sufi, A. 2012. The effects of fiscal stimulus: Evidence from the 2009 cash for clunkers program. *The Quarterly Journal of Economics* 127(3): 1107-1142.

Miguel, E. 2021. Evidence on Research Transparency in Economics. *Journal of Economic Perspectives* 35(3): 193-214.

Oak Ridge National Laboratory. 2015. Weatherization Works—Summary of Findings from the Retrospective Evaluation of the U.S. DOE's Weatherization Assistance Program. Washington, DC: DOE.

Office of Management and Budget (OMB). 2003. Circular A-4: Regulatory Analysis. September 17, 2003.

OMB. 2018. *2017 Report to Congress on the Benefits and Costs of Federal Regulations and Agency Compliance with the Unfunded Mandates Reform Act*.

OMB. 2019. Phase 1 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Learning Agendas, Personnel, and Planning Guidance. OMB Memorandum M-19-23.

OMB. 2020. Phase 4 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Program Evaluation Standards and Practices. OMB Memorandum M-19-23.

OMB. 2021a. Evidence-Based Policymaking: Learning Agendas and Annual Evaluation Plans. OMB Memorandum M-21-27.

OMB. 2021b. Interim Implementation Guidance for the Justice40 Initiative. OMB Memorandum M-21-28.

Popp, D., Vona, F., Marin, G., & Chen, Z. 2020. The employment impact of green fiscal push: evidence from the American Recovery Act. NBER WP 27321.

Robinson, L. A., Hammitt, J. K., & J Zeckhauser, R. 2016. Attention to distribution in US regulatory analyses. *Review of Environmental Economics and Policy* 10(2): 308-328.

Sunstein, Cass R. 2011. Empirically Informed Regulation. *University of Chicago Law Review* 78: 1349-1429.

Yagan, D. 2021. Using Evaluation to Advance Our Priorities. Blogpost at Evaluation.gov, URL: **https://www.evaluation.gov/2020-08-25-advance-our-priorities/**.