

Tracking and Evaluation of RD&D Programs at DOE: Concept Paper

Lucie Bioret and Alan Krupnick, Resources for the Future

May 16, 2023

With implementation of the Infrastructure Investment and Jobs Act (IIJA) and the Inflation Reduction Act (IRA), the Department of Energy (DOE) is to play a central role in undergirding the technological innovations needed to reach the Administration's net zero CO₂ emissions goal. Yet, to fulfil this mission it needs additional capacity in several areas. One area, which was the topic of our previous workshop, concerns how best to pick winners for the demonstration projects it will be funding. Another area, the topic of this workshop, is the development of systems for tracking and evaluating RD&D programs it funds, including but not limited to demonstration programs. Such a system is important for two reasons: it permits DOE and other interested parties to evaluate the success of programs in stimulating and, ultimately, commercializing technologies; and it provides input to the agency for adaptive learning using the results of program evaluations to improve its guidance to applicants (in the Funding Opportunity Announcements or FOAs), its decision protocols, and its project evaluations.

These key pieces of legislation are not the only motivation for the workshop. The Evidence Act from 2018 aims at modernizing federal government data collection and management processes to better inform policy decisions. It requires agencies to assess their current evaluation practices and create a plan to develop their evidence-building activities. Under the Biden administration, the Office of Management and Budget (OMB) is charged with improving evidence-building activities and evaluation by providing guidance and resources to agencies and by engaging with Evaluation Officers. At DOE, evaluation efforts and evidence-building activities fall under the responsibilities of each office¹, with Program Managers in charge of conducting evaluations while following the Program Evaluation standards².

This short paper is meant to prime workshop attendees for the event, which focuses on three topics: the current state of program evaluation at DOE and in other relevant agencies, the characteristics of robust evaluations methods and their associated metrics, and finally, the institutionalization of program evaluation and of supporting data collection within DOE. This paper will form the basis of a White Paper to be released shortly after the workshop that will incorporate information and ideas offered at the workshop by experts in academia and government, and by our own research.

Before starting out, we need to be clear about what an evaluation is and contrast it to an analysis. DOE refers to program evaluation as a “systematic assessment using quantitative and/or qualitative data and analysis methods to answer specific questions about current or past programs, with the intent to assess their effectiveness and efficiency.” Overall, the goal of program evaluation is to produce knowledge to improve programs. Program evaluations can assess how well a program performs compared to its goals and why it produces these results through impact evaluations, which are the main focus of our workshop. These try to identify the causal effects of a specific program on a range of outcomes, which often can involve using

¹ DOE, DOE Program and Functional Offices Evaluation/Evidence-Building Activities, FY 2024 Evaluation Plan, Learning Agenda, and Capacity Assessment, September 2022, <https://www.energy.gov/sites/default/files/2022-09/doe-evaluation-evidence-building-activities-fy24.pdf>

² OMB, Memorandum M-20-12, Phase 4 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Program Evaluation Standards and Practices, March 2020, <https://www.whitehouse.gov/wp-content/uploads/2020/03/M-20-12.pdf>

random controlled trials and other experimental approaches. Evaluations can also look at a program's progress by monitoring the implementation of an ongoing program with the goal of making process improvements³. In contrast, analyses at DOE are looking at the characteristics of funded technologies and how they are integrating into energy systems⁴, with the goal to produce knowledge for technology investment.

As shown above, recent legislation and the administration's commitment to evidence-based policy-making encourage agencies to develop program evaluation plans. We will discuss what this means for DOE and what lessons can be learned from other federal agencies such as the Department of Health and Human Services. We will then talk about DOE's evidence-building strategy and discuss evaluation efforts at the Energy Efficiency & Renewable Energy Office.

In the next session, we will discuss the characteristics of rigorous and robust impact evaluations, such as those comparing performance of grantees and ungranted applicants. We will also talk about alternative methods evaluators can use in the absence of this counterfactual information or due to other constraints.

Since we are focusing on RD&D programs, measuring innovation is a key part of the evaluation process, although it is difficult to quantify. We will discuss patent-related metrics that are often used as a measurable proxy for innovation to show knowledge creation and spillovers. Business-related metrics such as commercialization can also inform innovation diffusion, but data might be harder to procure.

The availability and collection of data is another main issue when evaluating RD&D programs. We will address in the workshop how data collection can be improved to reduce costs and deal with confidentiality issues. For instance, the data collection and reporting strategy could be embedded in the program design, so applicants understand their data provision responsibilities and confidentiality protections up front.

Designing programs with evaluation in mind is one way to strengthen the evaluation culture in DOE. We will address how institutional barriers can render evaluations difficult and how evidence building can be better incentivized through a deepening of an evaluation culture throughout the agency.

The following sections provide more details on each session of the workshop.

The Evaluation Landscape

Tracking and evaluation of programs is carried out by all agencies to various degrees, partly based on legislative directives, such as the Evidence Act or the legislation authorizing specific programs like the Small Business Innovation and Research (SBIR) Program⁵, and as a result of an agency's commitment to success and accountability.

DOE supports evidence-building activities and program evaluation to effectively reach their strategic goals. In DOE's FY2024 Evaluation Plan⁶ required under the 2018 Evidence Act and OMB guidelines, program evaluation is presented as key to managing a large portfolio of dissimilar programs and informing key

³ "EERE Program Evaluation," EERE, accessed May 8, 2023, <https://www.energy.gov/eere/analysis/eere-program-evaluation>

⁴ "EERE Strategic Analysis," EERE, accessed May 8, 2023, <https://www.energy.gov/eere/analysis/eere-energy-analysis>

⁵ Small Business Reauthorization Act of 2000, HR 5667. Pub. L. 106-554, 106th Congress, 2000, <https://www.congress.gov/bill/106th-congress/house-bill/5667>

⁶ DOE, DOE Program and Functional Offices Evaluation/Evidence-Building Activities, FY 2024 Evaluation Plan, Learning Agenda, and Capacity Assessment, September, 2022, <https://www.energy.gov/sites/default/files/2022-09/doe-evaluation-evidence-building-activities-fy24.pdf>

decisions on planning and budget. However, DOE does not have an all-of-agency evaluation strategy. Rather, their Plan focuses on processes and support, and delegates program evaluation responsibilities to functional offices and program managers. Although peer-reviews (a form of process evaluation) are becoming common practice in most Offices, we find that there are no systematic impact evaluations of RD&D programs to assess their effectiveness and efficiency, and no sufficient capacity to conduct them. And, as new Offices are created, it is important that programs are managed with future evaluations in mind.

EERE is thought to have a relatively good record for tracking and evaluating projects and programs among federal agencies, and especially technology agencies, according to contacted members of the evaluation community in other agencies. According to EERE evaluation requirements,⁷ programs and key projects must be peer-reviewed every two years by an independent panel of experts. EERE also encourages Technology Offices to conduct impact evaluations assessing causal effects of programs and outcomes against planned goals, although not in a systematic way and with a varied track record across Technology Offices. All these efforts are discussed in monthly Community of Practice meetings attended by tens of participants from across DOE.

Regarding evidence-building activities, efforts are on-going to track programs' outcomes across EERE with a database on technology commercialization developed by the Pacific Northwest National Lab for EERE. This database documents the time to market of EERE-funded technologies by soliciting projects' points of contact to collect data from companies' websites and scientific publications after projects' completion. Unfortunately, this effort, while long-standing, is not systematic, is very time-consuming, and can suffer from a low-response rate. In addition, it does not follow losing applicants, which is important for controlling for all the non-programmatic reasons for projects' successes and failures.

Other innovation outcomes, such as patents, are also targets of data collection efforts. In the context of a study on EERE-funded patents, researchers constructed a database containing all DOE grantees' patents (e.g. DOEPatents Database and iEdison Database). However, these data require lengthy processing and verification to link patents with specific programs and funding within DOE.

Although no other agency is a perfect analogue to DOE's activities, evaluation activities happening in other agencies under the Biden administration "all-of-government" evidence-based policy-making plan might be relevant for clean energy technology innovation programs. For instance, evidence-building activities are well integrated in some agencies like the U.S. Department of Health and Human Services, where there is an overarching evaluation strategy across programs carried out by independent and different offices.

Evaluating Innovation: Methods and Metrics

As in any evaluation, one needs to (i) define questions the evaluation is going to answer about the program (ii) define the unit of observation – in our case the project, (iii) collect data on the outcomes of interest to measure innovation success, such as technological uptake by purchasing firms, or on imperfect surrogates, such as the number of patents, (iv) consider the characteristics of the grants, such as size of award, timing, and other elements of the FOAs, and (v) consider the grantee's characteristics, such as revenues, profits, number of employees, private capital raised, etc.

If one observes such data for grantees and a similar comparison group, such as selected applicants that did not get funded, then it is possible to set up a quasi-experimental evaluation to quantify the causal impact of

⁷ EERE Evaluation Requirements, consulted on 5/8/2023, <https://www.energy.gov/eere/analysis/eere-evaluation-requirements>

the program (see Howell (2017) in Box 1). But other approaches are possible. For instance, when information on losers is not available, explaining the variations in outcomes among grantees could also help assess program success, be it in the selection process or the type of support provided to the grantees. Most impact evaluations conducted by EERE rely on mixed methods, including using expert elicitation to qualify the impact of RD&D funding on innovation for a specific technology compared to a counterfactual scenario. This workshop will mainly focus on the most rigorous and robust evaluation methods that rely on econometric analysis.

Since the workshop is focusing on RD&D programs, we expect to devote significant attention to metrics of success that quantify innovation. Regarding DOE RD&D programs, these can be classified in three categories; short term metrics derived from project performance in peer reviews such as technological advancement; medium term metrics, which are intermediate outcomes such as patent-related metrics to estimate knowledge creation, or business performance related metrics (private funding raised, firm closure); and long term metrics that spring from real world outcomes of technological innovation such as energy savings, CO₂ reductions, and market penetration (through sales) of the technologies targeted by grants. All are relevant although not all are necessary for a useful evaluation. Besides, not all metrics are easily measurable due to data availability constraints, among many reasons.

Box 1. Small Business Innovation and Research Program Evaluation (Howell, 2017)

In the context of a Congress-required evaluation by the National Academy of Sciences (NAS), a number of studies have been conducted on the SBIR program, such as Howell (2017) included in the follow up NAS study focusing on DOE's SBIR program (2020).

The SBIR program is a cross-agency grant program designed to stimulate innovations from small businesses and help commercialize them. DOE has participated in this program since 1982 and releases a multi-topic FOA twice a year to select new awardees. The Small Businesses Administration centrally collects the amount and names of recipients of all attributed awards. Since recipients are mainly small firms, finding data about innovation outcomes such as their patenting and business activities is fairly easy through public and private sources.

Howell's paper exploits the ranking of SBIR applications to compare the innovation and business performances of startups above and below the award cut-off to determine the program's impacts. The innovation performance is measured by the patent count associated with each start-up and business performance includes the amount of venture capital raised, firm acquisitions, and firm survival. Patent data comes from DOE internal databases and public sources while business-related data comes from proprietary databases. The ranking of grant applications, although crucial in the design of this evaluation, is not publicly available for DOE's SBIR nor for other competitive grant programs, and it can be very difficult to get permission to gain access to program data (NAS study, 2020).

In conclusion, Howell finds evidence of the innovation benefits of the program with the Phase 1 award increasing a firm's subsequent cite-weighted patents by at least 30 percent relative to non-awarded applicants, and its chances of receiving venture capital were more than 10 percent greater than those of non-awardees.

Perspectives on Institutionalization and Evaluation Culture

The best laid plans can come to naught unless there is a strong commitment from the top and middle management of an institution to see those plans through. This generalization is true of program evaluation. How to best build this evaluation culture is to be discussed at the workshop. The Evidence Act is a good start, as it provides the highest governance level of endorsement for systematic evaluations and evidence building activities. This commitment is reaffirmed in the current administration's budget proposal for 2024⁸, although no specific budget is attributed to build evaluation capacity in DOE where a large share of IJA and IRA money will be spent. Besides, while OMB is supporting the administration's goal, its authority is limited so its guidance

⁸ OMB, Analytical Perspectives: Budget of the U.S. Government FY2024– Chapter 12, March, 2023, https://www.whitehouse.gov/wp-content/uploads/2023/03/spec_fy2024.pdf

can only go so far in improving RD&D programs. Indeed, previous efforts to provide guidance and resources to regulatory agencies have not proven very effective for retrospective reviews⁹.

At its core, the barriers to building a tracking and evaluation system might be due to a lack of incentives to evaluate programs' effectiveness and efficiency. Indeed, if the results of evaluations are not used to recommend or justify policy actions, process changes, or programs improvements, investing resources in such assessments might be seen by program managers as a waste of their time and resources.

Another significant barrier could be the cost and effort needed to access data for evidence building activities. A pre-requisite to conduct RD&D program evaluation is access to data on applicants to measure outcomes and applicant characteristics. Yet, the FOA process is not designed to collect data for future evaluations. In addition, quantitative assessments are hampered by the lack of readily usable databases consolidating and organizing the data at the office or agency level, and the lack of available capacity to appropriately analyze the data and develop evidence-building activities.

When evaluating RD&D programs, another barrier might be the applicants' desire for secrecy. Many applicants don't want other applicants, shareholders, ratings agencies, and others to know of their plans. Some of this may be to avoid the embarrassment of losing, to limit the possibilities of giving away trade secrets, or giving an edge to competitors. Some may be due to sheer inertia favoring past processes. Also, asymmetric information can be beneficial for applicants that might prefer not to reveal their true performance to funders in the hope of receiving more funding in the future. Another possibility is that applicants might find reporting requirements burdensome, especially if they are a smaller organization.

To improve the data collection process and overcome barriers in program evaluation due to lack of data availability, several ideas can be advanced and further discussed during the workshop. First, FOAs could require applicants to report some information both before and after the award with some provision addressing Confidential Business Information. Second, current internal post-award tracking efforts could be improved and made available to evaluators and researchers (such as the iEdison database on federally funded patents and PNNL commercialization database). Third, data from other federal agencies could be used via special arrangements to make access easier. For instance, the Bureau of the Census already collects and stores a large quantity of high-quality plant-level data that could be leveraged by DOE and their evaluators through a partnership or another type of agreement.

⁹ Joseph E. Aldy, "Learning How to Build Back Better through Clean Energy Policy Evaluation," *Social Science Research Network*, August, 2022, <https://doi.org/10.2139/ssrn.4193992>