VOLUME

VALUE

CHAPTER

1

Introduction

Steven Sheingold

William Shrank

Background: Challenges and Opportunities

A comprehensive and dynamic evaluation should be an integral part of developing and implementing all major programs and policies. Planned and conducted properly, evaluations can provide important benefits to program administrators, policymakers, and the public. Evaluations are essential to understand the effect of programs and policies on the populations they serve, providing crucial evidence for the diffusion of innovations into practice. High-quality evaluation can provide the objective information needed to support the optimal prioritization and allocation of resources among existing programs and proposed new initiatives in the face of budgetary constraints. Additionally, evaluations can provide monitoring and feedback components that allow for continuous assessment, quality improvement, and mid-course corrections.

Conducting and using rigorous evaluations presents challenges, however. Evaluations can be costly and time consuming, and they can yield conflicting or inconclusive results. For these and other reasons, program proponents can see them—particularly in evaluations where people who are eligible for the program are assigned to a group that does not receive services—as draining away scarce resources that "should" be used to serve the maximum number of clients. In addition to these organizational concerns, applying appropriate evaluation methodologies to address a program's unique objectives can be extremely complex. This is particularly true for impact evaluations, which are intended to answer two major and related questions with a known degree of scientific credibility:

- Did target outcomes for the program improve, for whom did they improve, and under what circumstances?

- Can the findings demonstrate in a scientifically credible manner that the program, as opposed to other environmental factors, contributed significantly to the observed improvement (known as internal validity)?

Answering these questions can be especially challenging in cases where the intervention/new initiative cannot be evaluated with a built-in randomized controlled trial (RCT) design. An RCT design is considered the gold standard for impact evaluations, because it virtually guarantees an unbiased control group. Often, it is not practicable, and sometimes, it is even impossible. In such cases, alternative methods of achieving internal validity are required.

Another important issue is that a program evaluation design that is not fully integrated into the program's initial development and implementation frequently leads to inappropriate research designs and/or a lack of needed evaluation data—preventing the evaluation from yielding the type of authoritative results that can credibly establish the program's effectiveness or lack thereof. In cases where the evaluation has been designed and implemented separately from initial implementation, the observed findings—whether positive, negative, or inconclusive—can raise the following difficult question: Were the methods and data insufficient to

yield valid results—that is, to confidently establish whether the program was really effective or not?

While these challenges are very real, we believe that they can be minimized by a better understanding of evaluation methods and a clear vision of the utility of evaluation to promote learning, improvement, and rigorous outcome measurement. In our experience, obstacles to good evaluation, and the organizational concerns discussed above, arise from several factors: lack of understanding by program implementers of the variety of quasi-experimental and observational research designs that may be available when randomized trials are not possible; lack of understanding of the data flows required for implementing these methods; lack of understanding of relationships between monitoring, interim, or rapid formative evaluations and summative evaluation; and lack of familiarity with issues regarding translating evidence into practice or policy.

The age of "big data" also provides greatly expanded opportunities for overcoming past obstacles to evaluation. The increasing ability to collect, store, access, and analyze large quantities of data on a nearly real-time basis means that (1) multiple statistical methods may be available to improve the rigor of impact evaluations; (2) these methods can be applied on an ongoing basis; and (3) program monitoring can be seamlessly coordinated with the process and impact evaluation activities.

Evaluation and Health Care Delivery System Transformation

The delivery system transformation provisions of the Affordable Care Act (ACA) provided the impetus, indeed the necessity, of addressing the challenges and opportunities for strengthening the link between research and policy. ACA made the mandate clear—use innovation and experimentation to produce evidence that will help transform the health care system to deliver a higher quality of care at lower cost to Americans. Thus, developing methods to rigorously evaluate these programs and policies, and to do so expeditiously, has become critical.

The Center for Medicare and Medicaid Innovation (CMMI), which was authorized by the ACA, is charged with testing and evaluating innovative payment and service delivery models to reduce costs and improve quality of care in the broader, dynamic health care environment. The ACA provided $10 billion in CMMI funding from 2011 to 2019, combined with expanded authority to test innovative health care delivery system and payment models through Section 1115A of the Social Security Act (the Act).[1] The Act includes specific authority to expand program models if the evaluation finds, and the Centers for Medicare and Medicaid Services (CMS) Actuary certifies, that the model will either (1) reduce Medicare or Medicaid spending without reducing Medicare beneficiary access to care or (2) improve quality of care without increasing costs or reducing access. Once CMMI identifies the model, CMMI is responsible for development, implementation, and

evaluation to determine the effectiveness and feasibility of initiative expansion or scaling[2] to national program policy. This broad authority makes ensuring rigorous impact evaluation a critical part of CMMI's responsibility to rapidly identify and test desirable innovations, and use the resulting information to drive delivery system change.

CMMI faces some unique methodological issues in fulfilling these responsibilities. The health care delivery environment is extremely dynamic, presenting providers and other stakeholders with multiple, potentially conflicting, and complementary incentives for achieving key objectives. A range of public and private efforts may target reduction in hospital readmissions, for example, or hospital-acquired conditions. Indeed, CMMI itself may have several initiatives that address the same target outcome through different programs and incentives. Thus, attributing changes in outcomes to a specific CMMI payment or service delivery model—as opposed to all the other factors in the different program environments that might potentially influence the same outcome—presents a formidable methodological challenge. For example, CMMI models may be based on voluntary participation—which raises important concerns about selection bias. Why do some providers choose to participate, and others do not? The reasons are very unlikely to be random. Some see more opportunity for profit than others; for example, some may simply have too small (or too large) a market share to find it interesting to participate. Since an RCT design is not possible in a voluntary situation, a comparison group design is required to minimize the effect of the voluntary decision environment.

Finally, implementing initiatives within reasonable timeframes may require tradeoffs between program goals and methodological rigor. For example, it may not be possible to establish data collection methods conducive to constructing optimal comparison groups. It then becomes critical to make the best choice possible from quasi-experimental, or if necessary observational designs, that are still feasible under the circumstances.

Beyond these challenges to rigorous evaluation, CMMI maintains a focus on providing meaningful feedback to providers about performance during the implementation phase of a model—known as a formative (as opposed to a summative) evaluation. Only by rapidly comparing one site's performance to its own historical performance—as well as to the performance of other sites participating in the model and comparison sites that are not—can CMMI support the real-time learning and improvement essential to engender success. Moreover, CMMI must also collect contextual qualitative information about the program structure, leadership, and implementation, to understand strategies and features associated with success. This enables identification, harvesting, and dissemination of effective approaches. Understanding how to deliver rapid feedback without compromising the rigor of the summative evaluation is a key evaluation challenge.

More recent legislation has further emphasized the importance of these evaluation methods. Section 101(e)(1) of the "Medicare Access and CHIP Reauthorization Act of 2015" (MACRA) establishes the Physician Focused Payment Model (PFPM)

Technical Advisory Committee (TAC). The TAC is to provide comments and recommendations to the secretary of Health and Human Services on PFPMs submitted by stakeholders. The Committee must review submitted models, prepare comments and recommendations regarding whether such models meet criteria established by the Secretary, and submit those comments and recommendations to the Secretary. In order for the Committee to conduct its work in a valid, evidence based, and credible manner, it needs to have information to better understand the effectiveness of alternative payment models (APMs) on health care utilization, expenditures, and quality of care. Thus, they need to understand evaluation, results of ongoing evaluation of CMMI models, and how the various methodologies would be used for the PFPMs they propose.

These tasks are immediate and critical for delivery system change in a complex world. Such direct links between evidence and immediate policy decisions have already spurred considerable thinking about a broad range of evaluation methods and ways to directly translate evidence into practice and policy. Health services researchers and evaluators who have long hoped for a stronger link between research and policy may now be in a "be careful what you wish for" moment. The window is open for demonstrating that methodologically rigorous and policy relevant results can be produced within a timeframe that makes them most useful to decision makers.

The Global Context for Considering Evaluation Methods and Evidence-Based Decision Making

The ACA and CMMI might be considered as precipitating events for the new emphasis on these topics because of the statutory link between evaluation results and policy decisions. But delivery system change is hardly the only evaluation game in town. There is a much broader and practical context for expanding our thinking and increasing our knowledge about creating high-quality evidence for policy decision making. Indeed, the current economic and budgetary climate has created a greater need for improving our evaluation capabilities and the ability to integrate the evidence they provide into decision-making processes at all levels of government.[3] Most importantly, fiscal pressures demand that public programs and policies yield greater value, requiring the use of more evidence for policymaking. Such evidence is critical for developing programs, implementing them, assessing their effectiveness, and making needed adjustments.

The Congress recognized the importance of evidence-based policymaking to making government more effective by passing two laws that will have a substantial effect on the evaluation community. The first was the Evidence-Based Policymaking Commission Act of 2016 (Public Law 114–140, March 30, 2016). The law reflected a bipartisan call to improve the evidence available for making decisions about government programs and policies. The Commission was established to develop a strategy for increasing the availability and use of data to build

evidence about government programs, while protecting privacy and confidentiality. In September 2017, the Commission produced its final report and recommendations as required by the statute.[4]

In December 2018, the Congress passed Foundations for Evidence-Based Policymaking Act of 2017. The law will have substantial effects on government agencies with regard to formalizing evaluation planning and maintaining necessary data. It requires federal departments and agencies to submit annually to the Office of Management and Budget (OMB) and Congress a plan for identifying and addressing policy questions relevant to the programs, policies, and regulations of such departments and agencies. The plan must include (1) a list of policy-relevant questions for developing evidence to support policymaking, and (2) a list of data for facilitating the use of evidence in policymaking. The OMB shall consolidate such plans into a unified evidence-building plan.

The bill establishes an Interagency Council on Evaluation Policy to assist the OMB in supporting governmentwide evaluation activities and policies. The bill defines "evaluation" to mean an assessment using systematic data collection and analysis of one or more programs, policies, and organizations intended to assess their effectiveness and efficiency. Each department or agency shall designate a Chief Evaluation Officer to coordinate evidence-building activities and an official with statistical expertise to provide advice on statistical policy, techniques, and procedures. The OMB shall establish an Advisory Committee on Data for Evidence Building to advise on expanding access to and use of federal data for evidence building. Each agency shall (1) develop and maintain a comprehensive data inventory for all data assets created by or collected by the agency, and (2) designate a Chief Data Officer who shall be responsible for lifecycle data management and other specified functions.

These provisions do not, however, address how these data would be used to best inform the policymaking process. It will be necessary for program administrators and the research community to undertake the responsibility to develop rigorous evaluations pertinent to the unique questions raised by each program or policy area; conduct the evaluations in a timely manner; and learn how to translate the results in the most useful way to inform the decision makers in each situation.

This book will also be helpful to researchers in meeting these expanding responsibilities and to better understand the policy processes they hope to inform with their research. It will have particular value for readers who wish to concentrate on the most recent thinking on the methodological and translational issues *that are most key to informing program and policy decisions*. We recognize that a number of excellent primers and textbooks concerning evaluation are already available. In this text, we attempt to add to the knowledge base by comprehensively focusing on the translational and decision-making aspects of program evaluation, as well as providing a rigorous treatment of the methods issues. The evaluation, evidence, and decision-making topics discussed are applicable to most program and policy areas. Indeed, our illustrations and examples are drawn from a wide range of health care and non–health

care policy areas. We make it a point to provide illustrations (with both existing studies and original data analyses) to show how rigorous evaluation supports data-informed decision making. In the health program/policy area, the foremost goals are to improve health care services and reduce costs, without restricting health care access. Other program/policy areas have analogous goals.

Book's Intent

This book is intended to comprehensively address these important evaluation issues in a way that is rigorous enough to be useful for evaluators, while being easy to access by program administrators and policy decision makers whose primary focus is not methodological. This is a complex task to accomplish. Several chapters, by their nature, address issues of methodology. These chapters, primarily Chapters 5 and 6, include formal mathematical derivations. These are clearly marked, and readers not primarily interested in the methodological details can skip them. A number of chapters clarify important issues that need to be understood by both evaluators and decision makers *as they work together* to translate evaluation evidence into policy. These include analyzing and using interim results, harnessing alternative methods for analyzing and presenting data, and applying decision-making frameworks to empirical results for policy decision making.

The intent is to provide useful information for two basic situations. For those with *programs already under way*, the book can inform choices of methods, given the constraints imposed by existing program structure and data availability. For those who are *planning programs* and wish to *prospectively build in evaluation components*, the primer does two things: It provides guidance in the design of new programs to ensure they can be rigorously evaluated, and it highlights the types of data that would have to be generated for evaluation use.

This book also breaks new ground by addressing "rapid-cycle" evaluation in terms of state of the art methodologies, integration with monitoring and feedback requirements, and seamless transition to final impact evaluation. One clear source of skepticism about rigorous evaluations is that results often become available only when the time has passed for them to be useful. Thus, while the emergence of the term *rapid cycle* is associated with CMMI's unique responsibilities and authorities, the ability to produce credible results on a more rapid and continuous basis will benefit evaluations across a range of program areas.

DISCUSSION QUESTIONS

1. How can evaluators convince skeptical program proponents that conducting evaluations is a necessary practice?

2. How can the nearly real-time availability of data, and sometimes big data, change the way that evaluations are conducted? How can this impact the implementation of programs being evaluated?

3. What are the advantages and disadvantages of basing policy on evidence?

NOTES

1. The Social Security Act, 42 U.S.C., § 1115A(c)(1).

2. Section 1115A(c)(2)&(3) of the Act requires the Chief Actuary of CMS to certify such expansion would reduce (or would not result in any increase in) net program spending under applicable titles, and the Secretary determines that such expansion would not deny or limit the coverage or provision of benefits under applicable title for applicable individuals. The Secretary shall focus on models and demonstration projects that improve the quality of patient care and reduce spending.

3. Liebman, Jeffrey B., "Building on Recent Advances in Evidence-Based Policymaking," a paper jointly released by the Hamilton Project at Brookings Institution and Results for America, April 2013.

4. Report of the Commission on Evidence-Based Policymaking, *The Promise of Evidence-Based Policymaking*, September 2017.