Chapter 3:

Improving the Quality and Timeliness of Section 1115 Demonstration Evaluations

Improving the Quality and Timeliness of Section 1115 Demonstration Evaluations

Key Points

- Section 1115 of the Social Security Act provides the federal government with broad authority to waive certain Medicaid requirements to allow states to test demonstration projects likely to promote the objectives of the program.
- Under the statute, Section 1115 demonstrations must be evaluated, but state and federal administrations have historically focused on the flexibility offered under Section 1115 and placed limited emphasis on evaluation.
- Many evaluations have not generated findings that are timely or sufficiently rigorous to support decision making. The U.S. Government Accountability Office, MACPAC, and others have expressed concern regarding evaluation quality and how findings are used.
- To gather more specific information on issues in conducting evaluations and using findings, MACPAC convened an expert roundtable made up of state and federal Medicaid officials, evaluators of state demonstration programs, researchers, and other stakeholders. This chapter relies heavily on perspectives shared at the roundtable.
- The Centers for Medicare & Medicaid Services (CMS) has taken significant steps over the last five years to improve the quality of state-led evaluations, culminating with the 2019 release of new tools and guidance to help states develop strong evaluations.
- However, when planning and designing evaluations, states continue to struggle with methodological challenges, such as designation of comparison groups and availability of data. They also experience administrative challenges, such as constrained implementation timelines and budgets.
- States currently fund and direct the scope of evaluations, which has implications for evaluation independence and quality. States may be reluctant to devote time and resources to evaluation over program implementation, especially if doing so competes with political priorities.
- Given the importance of gathering evidence to inform decisions about the future of a demonstration policy, states should have an idea, before program implementation, of the measures and data sources they will use to assess whether the demonstration is making progress toward its objectives.
- Establishing appropriate standards of rigor and quality is difficult given state constraints. It may be appropriate to target standards related to content, rigor, and timing of evaluation deliverables according to demonstration type and scope.
- Achieving meaningful improvements in evaluations will take time, and will require CMS to continue focusing on these issues.
- At this time, MACPAC has not identified a need for further legislative or regulatory steps, but will continue to monitor how states and CMS carry out evaluations and use findings for decision making.

CHAPTER 3: Improving the Quality and Timeliness of Section 1115 Demonstration Evaluations

Section 1115 of the Social Security Act (the Act) provides the federal government with broad authority to waive federal Medicaid requirements to allow states to make changes to their Medicaid programs. Specifically, this authority allows the Secretary of the U.S. Department of Health and Human Services (the Secretary) to waive most of the requirements under Section 1902 of the Act to the extent necessary to enable a state to carry out an experimental, pilot, or demonstration project that the Secretary deems likely to assist in promoting the objectives of Medicaid (§ 1115 of the Act).[1] Section 1115 and its accompanying regulations require states to evaluate demonstrations approved under this authority (42 CFR 431.424).[2]

States have requested and received flexibility through Section 1115 authority to adopt a wide variety of innovations, including implementing alternative payment models and delivery systems, imposing additional eligibility criteria for certain beneficiary groups, providing new services to certain populations, and receiving federal matching funds for costs not otherwise matchable.[3] Federal administrations have also encouraged states to use Section 1115 authority to advance specific policy priorities. For example, the Centers for Medicare & Medicaid Services (CMS) recently approved Section 1115 demonstrations that allow states to adopt policies that have not been previously authorized, such as work and community engagement requirements as a condition of eligibility.

Robust evaluation findings about the effects of Section 1115 demonstrations can inform decision making at the state and federal levels. States can use findings from their own evaluations or those of other states to inform decisions such as applying for extensions or new demonstration authority. CMS can use evaluation findings to make approvals and develop new directions for federal Medicaid policy. Congress can use such information to make changes to the Medicaid statute. Historically, however, states and federal administrators have primarily focused on the flexibility offered under Section 1115 waivers, placing limited emphasis on evaluation and the role of a demonstration to produce evidence of the effects of new policies. Many demonstration evaluations have not generated findings that are timely or sufficiently rigorous to support decision making. Moreover, CMS has approved or extended many Section 1115 demonstrations with minimal evaluation evidence (GAO 2019, 2018). The U.S. Government Accountability Office (GAO), MACPAC, and others have expressed concern regarding evaluation quality and timeliness and how CMS uses findings to inform decisions.

CMS has taken significant steps over the last five years to improve the quality of state-led evaluations. Between 2017 and 2019, CMS released guidance outlining expectations for the content and research methods in evaluation design and reports, and a variety of other technical assistance resources (CMS 2019a). It also began including requirements for evaluation content and timing in the special terms and conditions (STCs) of each demonstration.[4] However, states continue to experience methodological and administrative challenges in carrying out strong evaluations, and the extent to which evaluations can or will be used to inform policy remains unclear.

This chapter begins by providing background information on Section 1115 demonstration authority in Medicaid and an overview of evaluation and monitoring requirements. It goes on to describe the concerns raised by GAO and CMS's recent efforts to improve evaluations.

The second portion of the chapter discusses the issues that remain for states and CMS when it comes to conducting evaluations and using

evidence; these include evaluation planning and funding, methodological challenges, timing issues, standards for evaluation quality, evidence needed to inform policy, and public comment and transparency. The discussion draws heavily on perspectives shared at a November 2019 expert roundtable convened by MACPAC to bring together state and federal officials, evaluators of several state demonstration programs, researchers, and other stakeholders.[5] The goals of this roundtable were to gather more specific information on the challenges states and CMS face in conducting timely, methodologically rigorous evaluations that can inform policy decisions; to solicit opinions on the appropriate balance of state flexibility and federal oversight; and to discuss potential steps that could be taken by states and the federal government to improve Section 1115 waiver evaluation processes.[6]

MACPAC is encouraged by the actions CMS has taken to help states conduct better evaluations. These actions appear to have also been well-received by states and evaluators. Evaluation guidance and evaluation-related STCs in place since 2017 will direct states receiving new or renewed demonstration approval but will not affect states with ongoing demonstrations. Achieving meaningful improvements in evaluation quality and usefulness will take time, and the effort requires CMS to remain vigilant in ensuring that states adhere to new expectations. At this time, MACPAC has not identified a need for further legislative or regulatory steps on this issue, but we will continue to monitor how states and CMS carry out evaluations and how these evaluations are used in decision making.

Use of Section 1115 Authority

Section 1115 predates the enactment of Medicaid as a vehicle for testing new approaches in a variety of federally funded programs and, in the early years of the program, was used infrequently for policy experimentation (MACPAC 2019). Its use has broadened over time: demonstrations authorized under Section 1115 over the last three decades have laid the groundwork for major Medicaid program changes. For example, the first Medicaid managed care programs were authorized under Section 1115 demonstration authority, as were the first programs offering Medicaid coverage of home- and community-based services (Rosenbaum 2017, Vladeck 1995). Several states used Section 1115 to expand coverage to low-income adults under age 65 not eligible on the basis of disability prior to the Medicaid expansion enacted in the Patient Protection and Affordable Care Act (ACA, P.L. 111-148, as amended) (Hinton et al. 2019, Holahan et al. 1995).

Section 1115 demonstrations are typically first approved for five years, and they can typically be extended for three or five years.[7] States apply for Section 1115 demonstrations by submitting a proposal to CMS; the proposal is the start of what can be a lengthy negotiation process between the two parties. The Secretary reviews each demonstration proposal to determine whether its stated objectives are aligned with those of the Medicaid program and whether proposed provisions and expenditures are consistent with federal policy (CMS 2019b). The Secretary has broad discretion to make such determinations and may do so in line with the administration's policy preferences. As such, Section 1115 has long been used by administrations as a mechanism to chart new Medicaid policy. Similarly, states use the flexibility afforded by Section 1115 to shape Medicaid policy to reflect their policy goals.

Currently authorized demonstrations

As of January 2020, there were 65 approved Section 1115 demonstrations underway in 47 states, with another 45 demonstration actions pending approval (including new demonstrations, amendments, and extensions) (CMS 2020). Although each of these demonstration programs has unique features, current demonstrations often do one or more of the following:

- require most or all Medicaid beneficiaries to enroll in managed care;

- adopt managed long-term services and supports (MLTSS) programs;
- implement delivery system reform programs;
- authorize federal spending for costs not otherwise matchable (e.g., uncompensated care pools);
- test alternative eligibility policies for low-income adults not eligible for Medicaid on the basis of disability (e.g., work and community engagement requirements or premiums as a condition of Medicaid eligibility);
- test strategies to address social determinants of health among certain populations or geographic areas;
- expand coverage to certain groups that would not otherwise be eligible for Medicaid in the state, such as individuals with HIV/AIDS or children with disabilities;
- expand access to certain benefits for individuals, such as those with substance use disorders (SUDs), or serious mental illness (SMI) and serious emotional disturbance (SED); or
- provide family planning benefits to certain populations.

Some of these demonstrations encompass most or all Medicaid beneficiaries or the entire state Medicaid program, while others target only a small subset of Medicaid beneficiaries or a discrete feature of the program. Some have been approved relatively recently, such as those implementing alternative policies for low-income adults not eligible on the basis of disability; others have been in place for decades, including managed care programs in several states.

Some of the policies included in these demonstrations can be implemented only through Section 1115 authority while others can be implemented under other authorities. For example, mandatory Medicaid managed care programs for most populations can be implemented through Section 1915(b) waiver authority or Section 1932 state plan authority. States using these authorities face more predictable approval application processes, and in the case of state plan authority, are not required to negotiate renewals or evaluate the program.[8] Many states, however, implement managed care under Section 1115 authority to show savings under the budget neutrality rules that can be used to finance other program changes (MACPAC 2018).[9] States may also prefer Section 1115 authority because it offers greater flexibility to limit certain services or policies to discrete populations. Additionally, states with long-standing managed care programs authorized under Section 1115 may lack administrative resources or capacity to reorganize their programs under a different authority.

Monitoring and Evaluation Requirements

Section 1115 demonstrations require both monitoring and evaluation; these are distinct activities with different purposes and timing. Monitoring provides ongoing updates on implementation and collects data on process and outcome measures, which may help states and CMS identify whether mid-course corrections are needed. Evaluations are completed later in the demonstration period or after the demonstration is complete; their purpose is to assess whether demonstrations have achieved their goals and to inform decisions about the future of the policy being tested. Although monitoring and evaluation are interrelated, the focus of this chapter is on evaluation.

Monitoring

All states must submit annual and quarterly monitoring reports describing the status of demonstration implementation and containing data on process and outcome measures.[10] According to federal regulations, several elements

must be included in annual monitoring reports, including early findings about the effect of the demonstration in meeting its objectives; a summary of grievances, appeals, and any feedback received from stakeholders; and information on various programmatic aspects of the demonstration (e.g., the number of people enrolled, legislative developments that affect the demonstration) (42 CFR 431.428). Quarterly reports typically contain implementation updates, a summary of press reports and issues arising during the quarter, and monitoring data. CMS requires states to collect and report monitoring data on specific metrics for certain demonstration types. For example, states with SUD demonstrations are required to report information on milestones and performance measures, such as trends related to assessment of need and qualification for SUD services, access to care, and the use of SUD-specific, evidence-based patient placement criteria (CMS 2019c).

States and CMS can use the information in monitoring reports to understand how implementation is affecting the program or its population and to make any needed mid-course adjustments. For example, monthly monitoring metrics reported in quarterly reports might show that beneficiaries in certain demographic subgroups are experiencing relatively high disenrollment rates. Such a finding might suggest the need to alter the specifications of the policy or provide additional support to those beneficiaries, or the need for additional analysis of the demonstration's effects by subgroup.

Evaluation

Currently, all states are required to submit a series of evaluation deliverables for each demonstration, including an evaluation design, an interim report, and a summative report. Requirements for these deliverables are set forth in federal regulation, the approved demonstration's STCs, and CMS evaluation design guidance.

Each deliverable must be submitted to CMS within a specified time frame:

- evaluation designs are due to CMS within 180 days following demonstration approval;
- interim evaluations are due with the demonstration renewal application or one year before demonstration expiration; and
- summative reports are due within 18 months of the end of the demonstration period.[11]

Evaluation designs specify the hypotheses being tested and describe the measures that will be used to assess progress toward the expected outcomes (42 CFR 431.424). They must include background information, research questions, methodology, and limitations and must identify the state's independent evaluator and include an evaluation budget (CMS 2019d).

CMS guidance requires reports to reflect the approved evaluation design. Both interim and summative evaluation reports must use quantitative methods whenever feasible and must minimize burden on beneficiaries and protect their privacy (42 CFR 431.424). Reports must also include results, conclusions about whether the demonstration met its goals, policy implications, details on interactions with other state initiatives, lessons learned, recommendations to policymakers and stakeholders, and a discussion of the study's limitations (CMS 2019e).

CMS reviews and provides comments on these evaluation deliverables and must approve them before they become final and are made publicly available. This provides the agency with several opportunities to guide the evaluation process.

Concerns about Evaluation Quality and Processes

GAO has repeatedly expressed concerns about the quality of evaluations, the timeliness of public release of evaluation findings, and the extent to which evaluations are used for policy decisions: In several studies published between 2007 and 2019,

GAO found that state-led evaluations have been limited by methodological shortcomings or selective reporting of outcomes, and that CMS has approved demonstration extensions despite evaluation results that are incomplete or inconclusive (GAO 2019, 2018, 2015, 2007).[12] Specific findings include:

- **Inconsistent application of evaluation requirements.** GAO has noted inconsistent application of evaluation requirements in several instances. For example, it found that CMS approved an extension application from Florida even though the state did not submit a required interim evaluation report as part of the application (GAO 2019). GAO also found that CMS deemed amendment applications from several states to be complete when they were missing information or when states had indicated that they were not planning to modify their evaluation designs, despite a requirement that states describe how they will do so when applying for amendments (GAO 2019).[13]
- **Significant methodological weaknesses.** GAO found that state-led evaluations often have significant methodological limitations that hamper their usefulness in informing decision making. Specifically, GAO noted that demonstrations in several states lacked adequate comparison groups or sufficient sample sizes and response rates for beneficiary surveys (GAO 2018).
- **Gaps in results.** In a review of several different demonstration types in a variety of states, GAO found that evaluations yielded few meaningful results. Specifically, it noted that several evaluations failed to address important hypotheses or report on key outcome measures for major aspects of the demonstrations. For example, under its delivery system reform incentive payment demonstration, Massachusetts was required to evaluate whether participating hospitals improved access to care, quality of care, and population health. However, the evaluation report included only descriptive information about the number and types of projects implemented by participating hospitals and did not report on effects or provide conclusions. Moreover, GAO noted a lack of evaluation results for many repeatedly renewed demonstrations that had never been subject to a final, comprehensive evaluation (GAO 2018).)[14]
- **Inadequate public comment processes.** GAO has repeatedly found that CMS approved demonstrations without adequate opportunity for, or consideration of, public input on their design and evaluation (GAO 2019, 2013, 2007). GAO has also observed that the extent to which CMS considered public comments in approving evaluation designs and evaluation components of STCs was unclear (GAO 2019). For example, it was unclear whether CMS considered public input in the approved evaluation design for the 2017 extension of Massachusetts' MassHealth demonstration, which did not include plans to examine the effects of its policy to discontinue provisional eligibility for most adults, despite concerns raised by many public commenters (GAO 2019). In other cases, CMS' feedback to states aligned with concerns raised in public comments. For example, CMS provided feedback on the evaluation design for Arkansas's work and community engagement demonstration, directing the state to address several concerns that were consistent with those raised during the public comment period (GAO 2019).

Efforts to Improve Evaluations

Evaluation requirements have evolved over the last two decades as Congress and CMS have made a number of changes to improve evaluation quality and processes (Table 3-1).

TABLE 3-1. Key Developments in Federal Policy for Evaluation of Demonstrations Approved under Section 1115 of the Social Security Act

Date	Action
1994	The Centers for Medicare & Medicaid Services (CMS) published policies and procedures for use in its review and approval of Section 1115 demonstrations, including high-level principles for evaluation.
2007	CMS issued a technical assistance guide for states, which highlighted basic principles and standards for the types of measures to use, comparison groups, and methods for distinguishing demonstration effects from other factors that could affect intended outcomes.
2010	Section 10201(i) of the ACA required the Secretary to establish a formal process for reviewing, approving, and conducting Section 1115 demonstration evaluations.
2012	In accordance with the ACA, CMS finalized regulations specifying monitoring and evaluation requirements. The regulations established a common set of reports that states must submit to CMS and make available on state websites, as well as minimum requirements for Section 1115 demonstration evaluations.
2017	CMS revised demonstration STCs to apply to demonstrations approved starting in 2017. The STCs describe common requirements and timing for evaluation designs and interim and summative evaluation reports for all demonstrations.[1]
2017	CMS released general evaluation design guidance setting forth its expectations for the format and content of evaluation designs, including required sections on background information, hypotheses and research questions, methodology, limitations, and information on the states' evaluator and evaluation budget.
2017	CMS issued an informational bulletin describing new strategies it planned to deploy in its review, approval, monitoring, and evaluation of Section 1115 demonstrations, including: • templates to streamline the initial application process for new demonstrations; • expedited approval of extensions of routine, successful, non-complex demonstrations for up to 10 years; and • fast-track approvals for established demonstrations that CMS found to have positive monitoring and evaluation results.
2019	CMS released additional guidance for developing evaluation designs in general and for specific demonstrations (such as work and community engagement demonstrations, demonstrations implementing premium requirements, and SUD and SMI/SED demonstrations).

Notes: Section 1115 is Section 1115 of the Social Security Act. ACA is the Patient Protection and Affordable Care Act (ACA, P.L. 111-148, as amended). The Secretary is the Secretary of the U.S. Department of Health and Human Services. STC is standard terms and conditions. SUD is substance use disorder. SMI is serious mental illness. SED is serious emotional disturbance.

[1] The current set of standard STCs regarding evaluation content and timing was implemented in 2017. Demonstrations approved earlier than 2017 have slightly different timing requirements for evaluation reports than those included in newer approvals. SUD-specific guidance was made available to states in 2018 but published by CMS to Medicaid.gov in 2019. CMS required states to follow this guidance on a state-by-state basis before it was released.

Sources: Bradley et al. 2019. CMS 2019a, 2017b, 2012, 2007. HCFA 1994.

Historically, such reforms have been geared toward promoting transparency and establishing expectations and consistent processes for evaluation content and timing. More recent reforms have emphasized improving quality and rigor.

Over the last five years, CMS has focused on strengthening state-led evaluations of Section 1115 demonstrations by providing guidance and technical assistance to states. Specifically, CMS has increased efforts to provide individualized feedback on draft evaluation deliverables, with attention to both compliance with requirements and technical rigor. CMS also published a new set of resources to help states in designing and executing their evaluations, including:

- a series of white papers that discuss common evaluation challenges (e.g., comparison group selection, best practices in causal inference);
- general evaluation design guidance, including requirements for the format and content of evaluation designs; and
- policy-specific guidance, with expectations for the components of evaluation designs for certain demonstration types (e.g., work and community engagement, premiums, SUD).[15]

The policy-specific guidance documents provide examples of logic models (or, in the case of the SUD guidance, a driver diagram) to help states think through a theory of change that incorporates hypotheses and expected outcomes. It also includes example design tables that suggest measures, data sources, and analytic approaches.

Participants at MACPAC's roundtable agreed that CMS's recent guidance has been important for establishing expectations for evaluations. Several participants pointed to the example logic models in policy-specific guidance as particularly helpful in encouraging states to consider their demonstration goals and anticipated outcomes.

CMS's recent efforts to raise evaluation standards will take time to yield meaningful progress. Since CMS began issuing policy-specific evaluation guidance, it has seen improvements in the quality of some states' draft evaluation designs, such as in the clarity of hypotheses and research questions.[16] However, it is too soon to know the full practical effects of the new guidance. Although CMS has approved and posted some evaluation designs that follow the new guidance to Medicaid.gov, no interim evaluation reports are yet available. Additionally, existing demonstrations will not be subject to new evaluation requirements until CMS incorporates them into the STCs for renewals.

Issues in Conducting Evaluations and Using Findings

There are many challenges to designing and carrying out strong evaluations. These include administrative challenges, such as limited evaluation budgets, lack of internal state expertise in research methods, and compressed implementation timelines. There are also methodological challenges (many of which are common to health services research in general), such as selecting comparison groups, obtaining reliable data, and separating effects of specific policies in multifaceted demonstrations. These challenges affect states to different degrees.

Evidence gathered through robust evaluation can help states and CMS make decisions about the future of the policies being tested. Historically, evaluations have not yielded findings that are rigorous or timely enough to be used for this purpose. But establishing appropriate standards of rigor and quality is difficult given the constraints states face. Setting evaluation schedules that produce timely findings is also difficult given data availability constraints. Even when robust and timely, findings may not be generalizable to other states, and they may not be sufficiently disseminated to inform broader policy discussions. We also note that decision making processes are influenced by a number of factors other than the evidence produced from specific evaluations,

including the desire to let states test new approaches and political and policy priorities of state and federal administrations.

Evaluation planning and funding

Although states must receive CMS approval of their evaluation designs, they have wide latitude in planning, budgeting, and procuring vendors to conduct the evaluation, and they vary in their approaches to doing so. The value proposition for investing time and resources in evaluations is not always clear to state legislators and executives. A disconnect between the statutory role of Section 1115 as a demonstration authority and state policymakers' use of Section 1115 as a mechanism for program flexibility may make state decision makers reluctant to invest in evaluation. This is often reflected in the evaluation budget, planning efforts, and overall quality of the evaluation, particularly if such investments are seen as competing with funds for the provision of health services. On the other hand, as one roundtable participant noted, some state Medicaid agencies may view evaluations as an opportunity to show state legislators a return on investment and persuade them to extend funding for the demonstration.

The state's role in directing and funding evaluations. The current arrangement, in which a state funds and directs the scope of evaluations, has implications for evaluation independence and quality. On the one hand, this arrangement allows those knowledgeable about the state's Medicaid program, its beneficiaries, and the available data sources to be closely involved with evaluation activities. On the other hand, it may limit the independence of evaluations, jeopardizing their quality. One risk of the arrangement is that it puts the state in charge of the budget rather than allowing evaluators or other entities to determine the level of funding required to conduct necessary evaluation activities. Political pressures or other state-specific circumstances may also influence how evaluators carry out evaluations and make decisions.

Evaluation budgeting. There are few requirements or guidelines for evaluation budgeting. Although recent CMS guidance has laid out expectations for evaluation format, content, and acceptable analytic methods, CMS has not provided explicit budget guidelines, including how budgets might vary based on demonstration characteristics, such as number of beneficiaries affected, complexity of the demonstration objectives, potential for adverse beneficiary consequences (e.g., disenrollment), and whether the demonstration authorizes policies that are relatively new and untested.

Lacking federal guidelines, states often determine evaluation budgets based on legislators' willingness to provide funds rather than on the cost of the necessary evaluation components. Other factors influencing budgets include the value policymakers place on evaluations (discussed above) and evaluation capacity and expertise among agency staff; states with relatively greater evaluation capacity are more likely to understand the level of funding needed to support strong evaluations. In some states, evaluation funds are not specifically allocated, meaning that any funds spent on evaluation reduce the amount available for Medicaid services or other administrative activities. This makes it more difficult for state agencies or evaluators to argue for larger evaluation budgets. Moreover, state budget cycles may not align with evaluation contracts; for example, a demonstration that is approved for five years may have an evaluation funded at a given level for the first year, but funding for subsequent evaluations may be subject to change from year to year during the state's budgeting process.

The appropriate level of evaluation spending is difficult to determine and will vary based on demonstration scope. Although states can use CMS guidance to help design their evaluations and can determine their budgets based on that design, conducting rigorous evaluations and adopting approaches recommended in the guidance, such as beneficiary surveys, may cost more than many states expect to spend. Additional CMS guidance and feedback could help states, for example,

guidance on typical costs for rigorous evaluation approaches or for evaluations of different types of demonstrations.

Roundtable participants noted that CMS could convey its commitment to improving evaluations by taking action such as increasing federal funding for evaluations or providing additional feedback and guidance on budgeting. Mechanisms for increasing federal investment in evaluations could include funding the full cost of evaluations or raising the federal matching rate for evaluations from the standard 50 percent administrative rate. Additionally, using a broader interpretation of the regulations governing the enhanced matching rate for mechanized claims processing and information retrieval systems could allow states to access a 75 percent or 90 percent federal matching rate for at least some evaluation activities.[17]

Early evaluation considerations. Efforts to consider evaluation early in the waiver application and implementation process may yield stronger evaluations. States typically begin evaluation planning after demonstrations are approved. However, discussing evaluations earlier—even before demonstration approval—could help states and CMS settle on demonstration designs that lend themselves to strong evaluation and give evaluators more time to design rigorous evaluation approaches. For example, evaluators can help states prioritize research questions and determine cost-effective ways to address them, assess needs for baseline data prior to implementing the demonstration, and help create an in-state comparison group by randomizing assignment to the demonstration or phasing in implementation.

CMS has recently begun encouraging states to involve evaluators as early in the process as possible; however, in many cases, states do not want to procure an evaluator until they have been granted approval for the demonstration or, in some cases, the evaluation design. Some roundtable participants noted that CMS could require states to demonstrate progress toward procuring an evaluator within a defined period of time after demonstration approval (e.g., by having a contract in place or identifying an evaluation design consultant or other partner). One roundtable participant suggested that CMS set up a contracting vehicle to allow states to access technical assistance resources from evaluation experts earlier in the process. Another roundtable participant suggested changing the Section 1115 demonstration application template to encourage states to define more explicitly what they are seeking to demonstrate and what hypotheses they would like to test.

Methodological challenges

There are methodological challenges in designing and carrying out Section 1115 evaluations, many of which are common in health services research and public program evaluation more generally, such as selecting comparison groups and obtaining reliable data. States and evaluators could address some of these challenges with advanced planning; in other cases, additional investment may be needed.

Comparison groups. Comparison groups are one of the most challenging methodological issues for evaluation. Absent a comparison group, it is difficult to understand whether changes in outcomes are due to demonstration policies or other factors. States have several options for selecting or constructing comparison groups, including in-state comparison groups consisting of Medicaid beneficiaries who are not subject to demonstration policies or other state comparison groups consisting of Medicaid beneficiaries in a similar state that does not have the same Section 1115 demonstration policies.

Both approaches have strengths and weaknesses. Using another state's Medicaid population as a comparison group may be appealing when comparable in-state populations are unavailable, but it may be difficult to find a comparable state. Data use agreements must be established in order to share individual-level administrative data across states; this can be difficult, especially if the comparison state has little incentive to participate.

Processing another state's data can be resource intensive because states use different data formats and file structures. In some cases, federal Medicaid data sets can be used for cross-state comparisons, although Transformed Medicaid Statistical Information System (T-MSIS) data have not been available until recently, and it is not yet clear if they can be used for evaluation.[18] One roundtable participant noted that there are efforts underway to enable the sharing of aggregated administrative data through distributed research networks such as the Medicaid Outcomes Distributed Research Network (MODRN). However, use of such data networks in Medicaid is not widespread.[19]

States can also use phased or randomized implementation strategies to construct comparison groups. These approaches do not require cooperation from other states and do not involve costs associated with sharing data, but still require careful planning and execution. States face several challenges when undertaking such advance planning, including coming into conflict with state priorities regarding implementation, and balancing those priorities against what is desirable or practical (e.g., the desire to provide new SUD services to the entire eligible population rather than a subset of individuals).

Alternatively, states can use analytic approaches that do not require comparison groups, such as interrupted time series analysis. Such approaches also require advance planning because they require many pre-period observations. They also may not be possible for states whose demonstrations have been in place for a long time and for which pre-period data are unavailable or outdated. Although such approaches cannot establish causal inference with the same level of rigor as approaches that use comparison groups, they can still produce useful information if properly designed (Bradley et al. 2019). However, planning and execution of these strategies may require staff expertise that Medicaid agencies do not have.

Obtaining data to examine particular demonstration outcomes. Data availability has long been a challenge for Medicaid research, including research on Section 1115 demonstrations. States and evaluators often lack the necessary data to address specific hypotheses. For example, Medicaid administrative data cannot be used to examine the effects of demonstration programs seeking to transition beneficiaries to commercial health insurance or outcomes that occur after leaving Medicaid. Additionally, Medicaid administrative data cannot measure many important outcomes of demonstrations. For example, in demonstrations implementing MLTSS, administrative data cannot provide insight regarding the extent to which the services and supports provided by managed care plans meet the needs and preferences of those receiving services, enhance community inclusion, and improve quality of life.

In some cases, administrative data from sources other than Medicaid can be used to examine such outcomes. For example, in a work and community engagement demonstration, unemployment insurance filings, tax returns, information in all-payer claims databases, or other data can be used to observe long-term outcomes among former beneficiaries or to track beneficiaries as they cycle on and off Medicaid. However, such data may be difficult to link to Medicaid data and may have other limitations.[20]

Beneficiary surveys are a key data source: they can assess beneficiary understanding of demonstration rules and incentives, help Medicaid programs connect with their beneficiaries, and generate evidence about beneficiary experience. However, such surveys can be challenging and expensive to administer, particularly those that follow beneficiaries over time to observe long-term outcomes. Due to low response rates, it can be difficult to achieve sufficient sample sizes for statistically sound analyses. National household surveys may include some data of interest for Medicaid beneficiaries, but they also present sample-size limitations and may not collect information on the specific population categories or policies that are relevant for demonstration evaluations. Other strategies

to gain insight on beneficiary behavior or experience include focus groups or systematic stakeholder interviews; these can yield qualitative information on beneficiary experience, but do not typically yield quantitative data that can be used to test all relevant hypotheses.

Estimating effects of specific policies in multifaceted demonstrations. Many states implement multiple policies through Section 1115 demonstrations that are intended to influence the same set of outcomes simultaneously. In other cases, a state's Section 1115 demonstration may be one of several concurrent initiatives that affect the Medicaid program and its beneficiaries—for example, a demonstration designed to provide services for beneficiaries in need of SUD treatment could be occurring alongside initiatives funded by other federal, state, or non-governmental programs with the same goal. This also makes it difficult to isolate the effects of one policy.

There are methodological strategies for disentangling the effects of multiple demonstration policies, such as randomization or sequential implementation of individual policies (Bradley et al. 2019; Reschovsky and Bradley 2019). However, these require advance planning and specialized expertise that states or even evaluators may lack.

Given these challenges and to help isolate the effects of specific policies, CMS has begun to encourage states to develop logic models for each policy and to focus on measuring outcomes that are likely to be affected by a single policy(CMS 2019d).

Timing

Timing requirements for evaluation deliverables vary by state and are generally linked to demonstration approval or expiration. These requirements, as currently structured, may contribute to the difficulty of conducting evaluations that can be used to inform policy.

Timing of evaluation design relative to demonstration implementation. It can take eight months or more after CMS approves a demonstration for the state to draft an evaluation design, obtain CMS comments, and gain approval (Bradley et al. 2019). Although CMS encourages states to plan for evaluation early, it does not currently require states to have an approved evaluation design prior to implementation.

States that move ahead without an approved evaluation design limit their options for robust analytic approaches. For example, experimental designs, the strongest method for public program evaluation, require randomized assignment into the demonstration, which must take place prior to implementation. Rigorous quasi-experimental designs may require the collection of some baseline data or a phased implementation in order to create a comparison group (Reschovsky and Bradley 2019). Under any approach, states should have an idea at the outset of the measures and data sources they will use to assess whether the demonstration is making progress toward its objectives.

This issue was highlighted by a recent experience in Arkansas, where the state received approval for community engagement requirements (through a demonstration amendment), implemented these policies, and then disenrolled beneficiaries for non-compliance before CMS had approved an evaluation design. In a letter to the Secretary, MACPAC expressed concern that without an approved evaluation design, Arkansas and CMS would not be able to interpret early experience with the demonstration, and that the short implementation time frame contributed to an "absence of sufficient measures and data to interpret early results and guide adjustments" (MACPAC 2018).[21]

On the other hand, it might not always be feasible to have an approved evaluation design plan in place prior to program implementation. States and CMS may prioritize program implementation and operation over evaluation. They may be bound by implementation timelines established by state law or by state procurement rules that slow down the process of obtaining an evaluator to work with on the evaluation design. There may also be pressure to move ahead quickly after approval, given the

five-year demonstration period, particularly in the case of demonstrations seeking to make broad, long-term program changes (e.g., delivery system reform). States may also be eager to begin providing access to new coverage or services for individuals (e.g., demonstrations that provide enhanced benefits for beneficiaries with SUD).

Timing of interim evaluation reports. States are required to submit interim evaluation reports to CMS with renewal applications or, if they are not seeking renewal, at least one year prior to demonstration expiration. The primary purpose of these reports is to inform CMS decisions about extension approvals. However, the timing of interim evaluation reports often results in short data collection periods, which limit the type of data that can be included and thus the usefulness of these reports. Depending on the length of the demonstration period and the details of the evaluation design, interim evaluations may be based on as little as a single year of demonstration experience. Data collected from the first year or two of a new demonstration, when the policies are not fully implemented and the operation of the program has not reached a steady state, might not be appropriate for inclusion in evaluations of policy outcomes or as the basis for awarding extensions.

In recent years, CMS has been more likely to approve five-year demonstrations. Although these longer approval periods provide more time for data collection and analysis before the interim report is due, there may still be significant data gaps. A three- to five-year data collection period is often insufficient to adequately assess the effects of a policy, especially if there have been delays in demonstration implementation or if changes have been made to the demonstration during implementation. When states make a mid-course change to the implementation approach or operational features of their demonstration, or if they pursue and receive approval for a demonstration amendment, evaluators may have to adjust the evaluation approach, and the timeframe in which data can be collected may be further reduced.

Currently, interim reports are generally intended to provide the same (or similar) information as the summative report, but at an earlier stage in the demonstration period. Noting the difficulty of collecting information on demonstration outcomes—some of which may be long-term outcomes—in short data collection periods, some roundtable participants cited examples of states working with CMS to clearly describe analyses that can be conducted in the interim versus the summative report.

Additionally, some roundtable participants suggested focusing interim reports on implementation. Information on process indicators (e.g., the share of providers participating in an intervention) or proximal outcomes (e.g., the share of beneficiaries who know about the incentive or requirement) could help indicate whether the demonstration has been implemented according to the design and provide information on how the demonstration is working. Evaluators could also collect and analyze qualitative data through key informant interviews and focus groups and assessments of program documentation. Information gathered through implementation research can help evaluators design analyses of outcome measures and can help states, CMS, and other stakeholders interpret findings on demonstration outcomes. This could improve the interpretation of findings from summative evaluation reports. Still, to serve their purpose of informing renewal decisions, interim evaluation reports would need to include some interim findings beyond implementation information.

Timing of summative reports. Summative evaluation reports, which are based on more years of data than interim reports, are due to CMS 18 months after the expiration of a demonstration cycle. This means that they are not available until after CMS decides to extend or renew the demonstration, which must generally happen by the end of a waiver cycle.[22] Findings from summative reports could inform future extensions or amendments, and may be of use to other states considering similar policies or to federal Medicaid policy deliberations more broadly. It

is important to note, however, that some long-term outcomes may not occur or be measurable within a five-year demonstration period.

Standards for evaluation quality

In recent evaluation design guidance, CMS clarified its expectations for the format and content of evaluations, including the hypotheses, measures, data sources, and analytic methods that would most likely produce strong evaluation findings. But because states are not required to adopt these approaches, CMS has begun to use STCs to describe hypotheses that states must articulate in order to test key demonstration policies on the outcomes of interest.[23] CMS also uses the guidance as a framework for reviewing states' designs and collaborating with them on improvements before approving revisions. Even so, there are no specific requirements for states to use certain methodological features. For example, CMS has not established standards for when specific components, such as comparison groups or beneficiary surveys, are necessary.

The wide variation in demonstration type and scope make it challenging to establish standards that would apply across all evaluations.[24] One possible approach would be to target standards and requirements related to content, rigor, and timing of evaluation deliverables according to demonstration type and scope. For example, roundtable participants and others in the policy community have raised the idea of categorizing demonstrations so that CMS can apply different standards and requirements to demonstrations of different types. Participants discussed several criteria that could be used, including risk of beneficiary harm, whether the policy being tested is a novel approach, the strength of the evidence for the policy, the likelihood of replication in other states, and the level of federal investment involved.[25]

Such an approach could require more rigorous evaluation features, such as randomized control groups and beneficiary surveys for demonstrations that pose high risk to beneficiaries (e.g., disenrollment for failure to comply with work and community engagement requirements) or involve a considerable federal investment (e.g., delivery system reform incentive payment programs). However, creating a system to categorize demonstrations would be difficult given different perspectives among decision makers and stakeholders about what constitutes risk or otherwise merits a higher standard of scrutiny.

Roundtable participants noted that greater collaboration between evaluators would be helpful for improving evaluations and establishing collective standards of rigor. For example, CMS could facilitate opportunities for states and evaluators to collaborate to improve skills, share lessons learned regarding demonstration evaluation, and distribute sample evaluation designs or evaluation requests for proposal across states.

Evidence needed to inform policy

Given the purpose of Section 1115 to allow states to experiment with new or different approaches, the statute anticipates that evidence gathered from formal program evaluations will address whether demonstrations achieve their objectives and the objectives of the Medicaid program as effectively or more effectively than the approaches permitted under current law. Despite this expectation, there are several long-standing demonstrations and demonstration policies that have been repeatedly extended with minimal evaluation evidence. For example, many states have had waivers of retroactive eligibility and non-emergency medical transportation policies in place for years. These features have received minimal attention by evaluators, however, and the effects of waiving these provisions of statute have not been clearly demonstrated.

In other cases, evidence is available but decision makers might find it difficult to assess whether there is enough evidence to make broader decisions such as providing statutory authority for the policy or, conversely, determining that the policy should not be permitted. For example, premiums and

cost sharing policies are commonly incorporated into Section 1115 demonstrations and have been studied extensively; evaluations have shown that premiums discourage enrollment, that cost sharing often leads to individuals avoiding care (including needed care), and that incentives for behavior change are poorly understood by enrollees and typically do not lead to the desired changes (MACPAC 2015, KCMU 2013). States, however, continue to seek and receive Section 1115 authority to implement such changes. In some cases, this may be because the changes they seek are variations on previous approaches, making the effects uncertain. In other cases, it may be because the literature is not well-known to state and federal policymakers, or because the findings conflict with policy priorities.

Judging the strength of evidence is not a straightforward undertaking and may require in-depth assessment of evaluation methods and interpretation of findings. Although standards have been proposed for rating the strength of evidence from clinical interventions, such standards can be difficult to apply to program evaluations given their complexity and the importance of context (Lohr 2004, Rychetnik et al. 2002).

Moreover, it is important to note that even high-quality evaluations that produce strong evidence can be of limited use in informing policy. Findings from one state's demonstration are unlikely to be definitive. Because of state-specific circumstances, differences in implementation design, or other factors, evaluation findings from one demonstration program may not produce information that is useful to another state that is looking to implement a similar policy. Moreover, Medicaid demonstration evaluations are designed to assess effects on the Medicaid program and on its beneficiaries, but are not designed to capture the effects of the demonstration on other aspects of the health care system or safety net, which can be meaningful.

Public comment and transparency

Federal regulations require federal and state public notice and comment periods for demonstration applications, but there are few opportunities for the public to comment on evaluation designs or findings. Interim reports are made available for public comment as part of the state's demonstration renewal application. However, relatively narrow dissemination of evaluation products limits the public's ability to review and comment on findings and the extent to which findings are shared and used by researchers and policymakers who are not otherwise involved in the demonstration.

Consideration of public input. The federal public notice process for state waiver application materials offers an opportunity for the public to comment on the pending application in light of any interim evaluation findings.[26] Public comments often raise concerns over certain demonstration features and can also be used to identify areas of risk that could benefit from careful evaluation. Federal rules for waiver applications in 42 CFR 431.412 do not explicitly require states to describe how public comments should inform evaluation hypotheses, and it is unclear if and how public comments are used when designing evaluations. A GAO report issued in 2019 noted that the extent to which CMS considers areas of risk identified through public input in evaluation designs and evaluation components of STCs is unclear and inconsistent (GAO 2019). Two roundtable participants noted that their state agencies share public comments with evaluators to inform demonstration hypotheses and research questions. However, it is not clear if this is a common practice. Several roundtable participants noted that it can be difficult to gather helpful public comments and feedback, citing low attendance at post-award forums and the lengthy and technical nature of evaluation.

Dissemination of findings. States and CMS are currently required to publish evaluation findings to their websites, but findings are not typically disseminated more broadly. Roundtable participants discussed how wider dissemination of evaluation findings, through a greater variety of channels, including post-award forums, blog posts, academic journals, and conferences, could expand the reach of these findings. Distilling findings so that they are easier to read and are more understandable to

lay audiences (e.g., through one-page summaries of findings) may also increase awareness and elicit higher-quality public comments.

Endnotes

[1] Generally, the Secretary may not waive provisions except those specified in Section 1902 of the Act. For example, the provisions related to federal medical assistance percentages (FMAPs) that are specified in Section 1903 may not be waived.

[2] The focus of this chapter is on state-led evaluations. However, it is important to note that the Centers for Medicare & Medicaid Services (CMS) at times conducts federal evaluations. For example, CMS sponsored a national, cross-state evaluation of several different types of Section 1115 demonstrations, underway from September 2014 through fiscal year (FY) 2019. Beginning in September 2018, CMS is also sponsoring federal evaluations through meta-analyses of certain types of Medicaid Section 1115 demonstrations. Additionally, CMS has sponsored several other state-specific and cross-state evaluations.

[3] Under Section 1115 authority, states can apply savings generated from portions of their demonstrations to request federal matching funds for costs that are not otherwise matchable under the state plan, making the demonstration budget neutral (§ 1115(a)(2) of the Act). These expenditures have been used to finance the following: coverage expansions to populations that are not otherwise eligible for Medicaid; additional payments to providers, such as uncompensated care pools or delivery system reform incentive payments; and additional payments to states (MACPAC 2018).

[4] Each approved Section 1115 demonstration is subject to STCs. These are legally binding documents that include a description of the statutory requirements being waived, the parameters of those waivers, state requirements and deliverables, beneficiary protections that must be guaranteed, budget neutrality calculations, and other terms of the waiver.

[5] The roundtable was held at MACPAC's office on November 14, 2019, and included officials from CMS and GAO; state Medicaid officials from four states; evaluators of state Section 1115 demonstration programs who were not from the same states as the Medicaid officials in attendance; researchers; and other stakeholders.

[6] MACPAC contracted with Mathematica to organize and moderate the roundtable and to prepare a background paper for participants, a formal agenda, and a summary report of the discussion.

[7] Customarily, CMS has approved initial demonstrations for five years and renewed them for up to five years (MACPAC 2019). However, in some cases, CMS has approved demonstrations for shorter or longer periods (CMS 2017a).

[8] Medicaid managed care programs implemented through Section 1915(b) waiver authority are subject to independent assessment rather than evaluation. States must contract with an independent entity to assess waiver performance during the first two years of operation and following the first renewal period. Independent assessments must address beneficiary access to services, quality of care, and cost-effectiveness of the waiver (MACPAC 2018).

[9] States can apply savings generated from the managed care portions (and other portions) of their demonstrations to request federal matching funds for costs not otherwise matchable and offset any associated additional costs to comply with the long-standing CMS policy that demonstrations be budget neutral to the federal government. Although many states using Section 1115 authority could operate their managed care programs under Section 1915(b) authority, doing so would limit their ability to use managed care savings to support additional spending under Section 1115 expenditure authority (MACPAC 2018).

[10] For some demonstrations, states submit annual or semiannual monitoring reports rather than quarterly reports (e.g., Maine Medicaid's Section 1115 Health Care Reform Demonstration for Individuals with HIV/AIDS and Georgia's Planning for Healthy Babies demonstration).

[11] These timing requirements were implemented after a policy change included in CMS's 2017 guidance and evaluation-related STCs. This means that demonstrations that have not been renewed since the policy change are not subject to this evaluation schedule.

[12] We note that GAO's reports to date generally examine evaluations for demonstrations approved prior to CMS's new guidance and evaluation requirements.

[13] Regulations governing application procedures require states to include their evaluation activities and findings to date in their extension applications (42 CFR 431.412(c)(2)(vi)). Regulations governing evaluations for extensions require the state to submit an interim report as part of the extension request (42 CFR 431.424(d)(1)).

[14] Prior to 2017, CMS policy required final, comprehensive evaluation reports after the expiration of the demonstrations rather than at the end of each three- to five-year demonstration cycle (GAO 2018). CMS changed this policy in 2017; summative evaluations are now due at the end of each demonstration cycle.

[15] Policy-specific guidance is currently available for SUD demonstrations, demonstrations focused on serious mental illness and serious emotional disturbance, and Section 1115 eligibility and coverage policies including community engagement requirements, premiums, non-eligibility periods (lockouts), and retroactive eligibility waivers (CMS 2019a).

[16] CMS does not usually release its review comments on states' evaluation products, but CMS posted comments on the draft evaluation designs submitted by Arkansas and Indiana to Medicaid.gov in 2018. Both states submitted their drafts before the release of the new eligibility and coverage evaluation design guidance in 2019. Some of the issues identified in these publicly released comments, such as unclear hypotheses and inadequate analytic approaches, have been improved in other states' draft designs for eligibility and coverage evaluations since the guidance was released. However, improvements are not consistent across states (Bradley et al. 2019).

[17] Federal law provides a 90 percent federal matching rate to the design, development, and implementation of mechanized claims processing and information retrieval systems and 75 percent for maintenance and operations of these systems (§ 1903(a)(3)(A)–(B) of the Act). A wider regulatory interpretation could allow states to access a 75 percent or 90 percent federal matching rate for at least some evaluation activities, such as coding, data analysis, and system development. CMS officials raised this possibility at the roundtable meeting, noting that this idea has been discussed, but no concrete action has yet been taken. Additionally, it is important to note that such a change may necessitate additional federal oversight to ensure that the enhanced matching rate is allowed only for activities authorized under current guidance.

[18] CMS and states are actively working on improving the quality and availability of T-MSIS data. CMS and others have expressed hope that the T-MSIS can resolve some of the data issues with evaluations, including permitting cross-state comparisons (GAO 2018).

[19] As of November 2019, MODRN is comprised of 11 state-university partnerships using a common data structure. Its primary cross-state project is an assessment of opioid use disorder treatment quality and outcomes in Medicaid, although it may be adapted for use in future Medicaid research on other topics (Sheets and Kennedy 2019).

[20] For example, in its initial draft evaluation design plan for its work and community engagement demonstration, Arkansas proposed using tax returns and the state's all payer claims database to track the income and health insurance status of beneficiaries who left Medicaid. In feedback provided to the state, CMS noted that evaluators for Arkansas's previous demonstration period were unable to track premium assistance beneficiaries into exchange coverage, and that tax returns might be limited as a data source because people with very low incomes are not required to file taxes (CMS 2018).

[21] Until Arkansas's demonstration was vacated by the U.S. District Court for the District of Columbia in March 2019, CMS was working with the state to develop an adequate evaluation design plan.

[22] The current requirement to submit summative reports after each demonstration approval period was instituted in 2017. Previously, final reports were due at demonstration expiration or closure rather than at the end of the approval cycle. This meant that states renewing their demonstrations were effectively not required to submit summative reports. To remedy this problem, CMS began including STCs that required states to submit summative reports after each demonstration approval cycle.

[23] CMS specifies these STCs in accordance with the evaluation guidance.

[24] In other types of Section 1115 demonstrations, the federal government has included more specific requirements for evaluation methodologies. For example, the Administration for Children and Families (ACF) generally required states with Section 1115 demonstrations that made changes to the Aid to Families with Dependent Children (AFDC) program (e.g., welfare reform demonstrations) to use an experimental design with a randomized selection process. Further details, such as sample size, procedures for drawing the sample, and control processes for maintaining the integrity of the design, were negotiated between states and ACF. Like Medicaid demonstrations, AFDC demonstrations also varied by state, and their evaluation costs were shared evenly between the state and federal government (Harvey et al. 2000).

[25] In November 2017, CMS announced plans to make distinctions in the level of evaluation and monitoring required for different demonstrations. The guidance includes some broad criteria, including demonstrations that are long-standing, non-complex, and unchanged; have previously been rigorously evaluated and determined successful, without issues or concerns that would require more regular reporting; include a small number of enrollees (approximately 500 or less); have been operating smoothly without administrative changes; have been subject to only a minimal number of appeals and grievances; have no state issues with CMS 64 reporting or budget neutrality; or do not have a corrective action plan in place (CMS 2017b). Further details have not been released. However, following the release of this guidance, GAO issued a 2018 report recommending that CMS issue written criteria for when it will allow limited evaluation of a demonstration, including defining what it means for a demonstration to meet the various conditions identified (e.g., long-standing or non-complex) (GAO 2018).

[26] Federal regulations at 42 CFR 431.408 and 431.416 require both state and federal public notices for demonstration applications. The regulations outline specific content for public notices, including proposed demonstration policies and hypotheses to be tested.

References

Bradley, K., D. Lipson, and J. Self. 2019. Improving the quality and timeliness of Section 1115 demonstration evaluations. Paper prepared for the Medicaid and CHIP Payment and Access Commission as background material for participants in a roundtable meeting on November 14, 2019.

Centers for Medicare & Medicaid Services (CMS), U.S. Department of Health and Human Services. 2020. State waivers list. Baltimore, MD: CMS. https://www.medicaid.gov/medicaid/section-1115-demo/demonstration-and-waiver-list/index.html.

Centers for Medicare & Medicaid Services (CMS), U.S. Department of Health and Human Services. 2019a. 1115 demonstration state monitoring & evaluation resources. Baltimore, MD: CMS. https://www.medicaid.gov/medicaid/section-1115-demo/evaluation-reports/evaluation-designs-and-reports/index.html.

Centers for Medicare & Medicaid Services (CMS), U.S. Department of Health and Human Services. 2019b. About Section 1115 demonstrations. https://www.medicaid.gov/medicaid/section-1115-demo/about-1115/index.html.

Centers for Medicare & Medicaid Services (CMS), U.S. Department of Health and Human Services. 2019c. Monitoring metrics for Section 1115 demonstrations with SUD policies. Baltimore, MD: CMS. https://www.medicaid.gov/medicaid/section-1115-demonstrations/1115-demonstration-monitoring-evaluation/1115-demonstration-state-monitoring-evaluation-resources/index.html.

Centers for Medicare & Medicaid Services (CMS), U.S. Department of Health and Human Services. 2019d. Section 1115 demonstrations: Developing the evaluation design. Baltimore, MD: CMS. https://www.medicaid.gov/medicaid/section-1115-demonstrations/1115-demonstration-monitoring-evaluation/1115-demonstration-state-monitoring-evaluation-resources/index.html.

Centers for Medicare & Medicaid Services (CMS), U.S. Department of Health and Human Services. 2019e. Section 1115 demonstrations: Preparing the evaluation report. Baltimore, MD: CMS. https://www.medicaid.gov/medicaid/section-1115-demonstrations/1115-demonstration-monitoring-evaluation/1115-demonstration-state-monitoring-evaluation-resources/index.html.

Centers for Medicare & Medicaid Services (CMS), U.S. Department of Health and Human Services. 2018. Letter from Andrea Casart to Dawn Stehle regarding "CMS letter and feedback to Arkansas on draft evaluation design." November 1, 2018. https://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Waivers/1115/downloads/ar/Health-Care-Independence-Program-Private-Option/ar-works-feedback-eval-dsgn-20181101.pdf.

Centers for Medicare & Medicaid Services (CMS), U.S. Department of Health and Human Services. 2017a. CMS approved first 10-year Section 1115 demonstration extension. December 28, 2017, press release. Baltimore, MD: CMS. https://www.cms.gov/newsroom/press-releases/cms-approves-first-10-year-section-1115-demonstration-extension.

Centers for Medicare & Medicaid Services (CMS), U.S. Department of Health and Human Services. 2017b. CMCS informational bulletin regarding "Section 1115 demonstration process improvements." November 6, 2017. Baltimore, MD: CMS. https://www.medicaid.gov/federal-policy-guidance/downloads/cib110617.pdf.

Centers for Medicare & Medicaid Services (CMS), U.S. Department of Health and Human Services. 2012. Medicaid Program: Review and approval process for Section 1115 demonstrations. Final Rule. *Federal Register* 77, no. 38 (February 27): 11678–11700. https://www.federalregister.gov/documents/2012/02/27/2012-4354/medicaid-program-review-and-approval-process-for-section-1115-demonstrations.

Centers for Medicare & Medicaid Services. U.S. Department of Health and Human Services. 2007. *Evaluating demonstrations: A technical assistance guide for states*. Baltimore, MD: CMS.

Harvey, C., M.J. Camasso, and R. Jagannathan. 2000. Evaluating welfare reform waivers under Section 1115. *Journal of Economic Perspectives* 14, no. 4: 165–188. https://pubs.aeaweb.org/doi/pdf/10.1257/jep.14.4.165.

Health Care Financing Administration (HCFA), U.S. Department of Health and Human Services. 1994. Medicaid Program: Demonstration proposals pursuant to Section 1115(a) of the Social Security Act; Policies and procedures. Public notice. *Federal Register* 59, no. 186 (September 27). https://www.govinfo.gov/content/pkg/FR-1994-09-27/html/94-23960.htm.

Hinton, E., M. Musumeci, R. Rudowitz, et al. 2019. *Section 1115 Medicaid demonstration waivers: The current landscape of approved and pending waivers*. February 2019. Washington, DC: Kaiser Family Foundation. http://files.kff.org/attachment/Issue-Brief-Section-1115-Medicaid-Demonstration-Waivers-The-Current-Landscape-of-Approved-and-Pending-Waivers.

Holahan, J., T. Coughlin, L. Ku, et al. 1995. Insuring the poor through Section 1115 Medicaid waivers. *Health Affairs* 14, no. 1: 199–216. https://www.healthaffairs.org/doi/pdf/10.1377/hlthaff.14.1.199.

Kaiser Commission on Medicaid and the Uninsured (KCMU). 2013. *Premiums and cost-sharing in Medicaid: A review of research findings.* Washington, DC: KCMU. https://www.kff.org/wp-content/uploads/2013/02/8417-premiums-and-cost-sharing-in-medicaid.pdf.

Lohr, K. 2004. Rating the strength of scientific evidence: Relevance for quality improvement programs. *International Journal for Quality in Health Care* 16, no. 1: 9–18. https://academic.oup.com/intqhc/article/16/1/9/1820191.

Medicaid and CHIP Payment and Access Commission (MACPAC). 2019. Section 1115 research and demonstration waivers. Washington, DC: MACPAC. https://www.macpac.gov/subtopic/section-1115-research-and-demonstration-waivers/.

Medicaid and CHIP Payment and Access Commission (MACPAC). 2018. Chapter 1: Streamlining Medicaid managed care authority. In *Report to Congress on Medicaid and CHIP*. March 2018. Washington, DC: MACPAC. https://www.macpac.gov/wp-content/uploads/2018/03/Streamlining-Medicaid-Managed-Care-Authority.pdf.

Medicaid and CHIP Payment and Access Commission (MACPAC). 2015. *The effect of premiums and cost sharing on access and outcomes for low-income children*. July 2015. Washington, DC: MACPAC. https://www.macpac.gov/publication/the-effect-of-premiums-and-cost-sharing-on-access-and-outcomes-for-low-income-children/.

Reschovsky, J., and K. Bradley. 2019. *Planning Section 1115 demonstration implementation to enable strong evaluation designs*. March 2019. Washington, DC: Mathematica. https://www.medicaid.gov/medicaid/section-1115-demo/downloads/evaluation-reports/enable-strng-eval-dsgn.pdf.

Rosenbaum, S. 2017. Weakening Medicaid from within. *American Prospect*. October 19. https://prospect.org/power/weakening-medicaid-within/.

Rychetnik, L., M. Frommer, P. Hawe, and A. Shiell. 2002. Criteria for evaluating evidence on public health interventions. *Journal of Epidemiology and Community Health* 2002, 56: 119–127. https://jech.bmj.com/content/jech/56/2/119.full.pdf.

Sheets, L., and S. Kennedy. 2019. *An overview of the Medicaid Outcomes Distributed Research Network (MODRN)*. Washington, DC: Academy Health. https://www.academyhealth.org/publications/2019-11/overview-medicaid-outcomes-distributed-research-network.

U.S. Government Accountability Office (GAO). 2019. *Medicaid demonstrations: Approvals of major changes need increased transparency.* Report no. GAO-19-315. Washington DC: GAO. https://www.gao.gov/products/GAO-19-315.

U.S. Government Accountability Office (GAO). 2018. *Medicaid demonstrations: Evaluations yielded limited results, underscoring need for changes to federal policies and procedures*. Report no. GAO-18-220. Washington, DC: GAO. https://www.gao.gov/products/GAO-18-220.

U.S. Government Accountability Office (GAO). 2015. *Medicaid demonstrations: Approval criteria and documentation need to show how spending furthers Medicaid objectives*. Report no. GAO-15-239. Washington, DC: GAO. https://www.gao.gov/products/GAO-15-239.

U.S. Government Accountability Office (GAO). 2013. *Medicaid demonstration waivers: Approval process raises cost concerns and lacks transparency*. Report no. GAO-13-384. Washington, DC: GAO. https://www.gao.gov/products/GAO-13-384.

U.S. Government Accountability Office (GAO). 2007. *Medicaid demonstration waivers: Lack of opportunity for public input during federal approval process still a concern*. Report no. GAO-07-694R. Washington, DC: GAO. https://www.gao.gov/products/GAO-07-694R.

Vladeck, B. 1995. Medicaid 1115 demonstrations: Progress through partnership. *Health Affairs* 14, no. 1: 217–220. https://www.healthaffairs.org/doi/pdf/10.1377/hlthaff.14.1.217.