

EVALUATION DESIGN GUIDANCE FOR SECTION 1115 ELIGIBILITY AND COVERAGE DEMONSTRATIONS

Introduction

The purpose of this document is to support states in developing an evaluation design for eligibility and coverage demonstrations authorized through section 1115, and to communicate CMS expectations for the rigor of demonstration evaluations. This document should be used as a basis for states' discussions with their independent evaluators, and may also support evaluator procurement. It is not intended to replace the evaluation design process or to resolve all design-related issues that may arise. We encourage states to procure their evaluator to support the development of a robust draft evaluation design for CMS review; ideally, states should identify an evaluation team before implementation to consult on implementation plans that support robust research designs and plan early data collection. CMS is available to provide further technical assistance to states if requested.

CMS has approved several demonstrations with policies that affect Medicaid eligibility and coverage, including community engagement requirements, premiums or monthly beneficiary account contributions,¹ non-eligibility periods as a consequence for noncompliance with program requirements, and retroactive eligibility waivers. Specific evaluation design guidance for each of these policies is available as an appendix to this document; states with more than one eligibility and coverage policy should consult each relevant appendix to build their demonstration evaluation design. Note that states with more than one eligibility and coverage policy may not be able to address all recommended research questions in each appendix because it will not be possible to attribute observed effects to individual policies, as opposed to the demonstration as a whole. States should work with their evaluators to determine which research questions are most appropriate and feasible to address for individual demonstration policies.

CMS has also provided an appendix with a suggested approach to using information on cost impacts and other evaluation evidence to assess demonstration sustainability. The appendix provides specific suggestions for analyzing demonstration costs to the state, as well as general considerations for assessing sustainability. Because there is no single, direct measure of sustainability, states must make judgments that weigh evaluation evidence within the context of the stated objectives of their section 1115 demonstrations and relevant time horizons for policy and budgetary decisions.

¹ Beneficiary account policies may also involve specific incentives for certain health behaviors. This guidance does not include recommendations for assessing effects of specific health behaviors because such incentives vary widely across states. States and their evaluators may wish to consult "Best Practices in Causal Inference for Evaluations of Section 1115 Eligibility and Coverage Demonstrations" (<https://www.medicaid.gov/medicaid/section-1115-demo/downloads/evaluation-reports/causal-inference.pdf>), the evaluation plan for the 2014-2019 national evaluation of section 1115 demonstrations with beneficiary engagement policies (<https://www.medicaid.gov/medicaid/section-1115-demo/downloads/evaluation-reports/eval-plan-beneficiary-engagement-programs.pdf>), or other states' publicly available evaluation reports.

CMS evaluation policy

Consistent with the Administration's goal to hold states accountable for outcomes associated with their Medicaid programs, states with approved section 1115 eligibility and coverage demonstrations are required to conduct regular and robust monitoring and rigorous evaluation activities. CMS recommends an integrated approach to monitoring and evaluation, as monitoring data provide essential information on demonstration implementation, creating the context for evaluations and informing interpretation of results. Monitoring data can also be used as a data source for evaluation research questions focused on demonstration processes.

This document contains general guidance for developing the evaluation design for section 1115 eligibility and coverage demonstrations. Separately, CMS will provide guidance for monitoring approved demonstrations, and encourages states to share monitoring plans and data with contracted evaluators.

As part of the requirements described in the demonstration special terms and conditions (STCs), states are expected to submit to CMS an evaluation design representative of the state-specific demonstration within 180 days after approval. CMS must approve the evaluation design, and every effort should be made to follow the approved evaluation design when conducting analyses and developing interim and summative evaluation reports. However, the state may request, and CMS may agree to, changes in the methodology in appropriate circumstances.

Section 1115 Medicaid Demonstration Evaluations

Federal law specifically authorizes experimentation by state Medicaid programs through section 1115 of the Social Security Act. Under section 1115 provisions, states may apply for federal permission to implement and test new approaches to administering Medicaid programs that depart from existing federal rules yet are consistent with the overall goals of the program, likely to meet the objectives of Medicaid, and budget neutral to the federal government. Section 1115 also requires that federal regulations specify requirements for demonstration evaluations. 42 CFR 431.424 outlines general evaluation requirements and evaluation design components such as hypotheses, data sources, and comparison strategies. 42 CFR 431.412 outlines requirements for including initial evaluation plans and evaluation reports in demonstration applications and renewal applications, respectively. STCs for each state further specify CMS expectations for the contents and timing of evaluation designs and interim and summative evaluation reports.

How to use this technical assistance guide

This guidance is intended to support states in developing evaluation designs that will meet CMS expectations for rigorous state evaluations and comply with evaluation requirements in demonstration STCs. This guidance includes hypotheses, research questions, and evaluation approaches that will generate strong evaluation designs. Hypotheses and research questions for individual policies are contained in separate design appendices for each policy. Hypotheses for section 1115 community engagement demonstrations are consistent with State Medicaid Director letter 18-002, which communicates CMS's expectation that states test the effects of community engagement requirements on health, well-being, independence, and the sustainability of the Medicaid program. Recommended evaluation approaches are based on best practices in policy evaluation and recent state evaluations of eligibility and coverage policies authorized through section 1115.

The content included in an evaluation design submitted to CMS by each state will need to be customized to each state's specific demonstration. For example, states may wish to modify or add hypotheses or research questions to ensure alignment with the state's section 1115 demonstration policies. CMS recognizes that each state's demonstration is unique and therefore the evaluation design submitted to CMS should capture those unique elements. States should prioritize the information that will best provide evidence for the policies the state is testing under section 1115 authority, and will best inform decisions regarding future policy. States may find this guidance, and the monitoring and evaluation requirements in the demonstration STCs, helpful for specifying and allocating the resources needed to conduct a robust evaluation.

This guidance also provides a general framework with which the state and its evaluators can begin writing the evaluation design.

The format for the evaluation design should be as follows:

- A. General background information
- B. Evaluation questions and hypotheses
- C. Methodology
- D. Methodological limitations
- E. Attachments

These sections correspond to the general evaluation design guidance that CMS has frequently incorporated as part of STCs.² The remainder of this document provides guidance on how to complete each of these sections.

A. General background information

The state should include basic information about the demonstration, as follows.

1. The name of the demonstration, approval date of the demonstration, and period of time covered by the evaluation. If the demonstration approval period and the evaluation time period are not identical, the state should indicate the difference and provide a brief explanation (e.g., the state was approved for the demonstration in January 2018, but did not start implementation until April 2018).

2. The goals of the demonstration policies. This should be state-specific, and should include goals for each policy contained in the demonstration. Potential goals for individual eligibility and coverage policies are included in the design appendix for each policy. For example:

The purpose of the demonstration is to test whether requiring community engagement activities as a condition of eligibility:

- a. Leads to increased or sustained employment,
- b. Improves beneficiaries' socio-economic status,

² This document and related appendices focus on evaluation of eligibility and coverage policies and therefore contain slight differences from CMS's general guidance, "Section 1115 Demonstrations: Developing the Evaluation Design," which is also available at <https://www.medicaid.gov/medicaid/section-1115-demo/downloads/evaluation-reports/developing-the-evaluation-design.pdf>.

- c. Promotes beneficiary independence, and
- d. Improves health outcomes.

3. A brief description of the demonstration and the implementation plan. This can be a summary of information provided in the implementation plan. The summary should include information relevant to the evaluation design, such as implementation timing and any plans for staged roll-out of relevant policies by geographic area or eligibility group, which may create opportunities for within-state comparison groups.

4. Describe the population groups impacted by the demonstration. The state should include a description of adult eligibility groups that are subject to relevant policies, as well as those that are not subject to those policies and that may therefore serve as within-state comparison groups for the evaluation.

5. Other relevant contextual factors. The state should note other state-wide health care delivery or payment reform efforts that may affect either the demonstration or the evaluation or both.

B. Evaluation hypotheses and research questions

1. Logic model. A primary purpose of section 1115 demonstration evaluations is to determine whether demonstrations are achieving their stated goals for each demonstration policy. Logic models³ help states and their evaluators harmonize the goals of the evaluation with the goals of the policy. Logic models should depict expected short-term, intermediate, and long-term outcomes for each policy. Depicting the order and timing of expected policy outcomes is particularly important for eligibility and coverage policies because certain outcomes may not occur for a long time. Relatedly, because hypotheses for some policies are logically dependent (i.e., later outcomes are dependent on realizing earlier outcomes), states and their contracted evaluators should use evaluation results for short-term and intermediate outcomes to interpret results for later outcomes.

States should also include moderating factors and confounding (or contextual) variables in their logic models. Moderating factors are important preliminary outcomes that states should consider because they affect the relationship between the demonstration policy and one or more hypothesized outcomes; however, they are not themselves the policy goals. For example, beneficiary understanding of demonstration policies is a moderating factor that may affect—and should inform interpretation of—observed demonstration outcomes. Confounding or contextual variables may influence policy implementation or outcomes, and can bias evaluation results if the evaluation approach does not control for them. Beneficiaries' underlying health status is an example of a confounding variable that evaluators should control for in any regression model of

³ As noted in “Best Practices in Causal Inference for Evaluations of Section 1115 Eligibility and Coverage Demonstrations” (<https://www.medicaid.gov/medicaid/section-1115-demo/downloads/evaluation-reports/causal-inference.pdf>), both logic models and driver diagrams depict a theory of change that supports evaluation design. Driver diagrams typically focus on factors that must change in order to achieve a policy goal. The following link provides a helpful description of the process for developing a driver diagram (<https://innovation.cms.gov/files/x/hciatwoaimsdrrvs.pdf>). Logic models are emphasized here to focus on relationships between measurable policy outcomes. In practice, these diagrams can serve similar purposes in evaluation planning.

the effect of demonstration policies on health outcomes. Finally, states may also wish to include potential unintended or adverse consequences. An example logic model for individual eligibility and coverage policies is included in the design appendix for each policy.

2. Hypotheses and research questions. Evaluation documents must include a discussion of the hypotheses that the state intends to test, along with research questions that will address each hypothesis. Hypotheses should correspond to the policy goals for each eligibility and coverage policy and should be clearly written to state the direction of the expected change, although evaluators should use two-sided hypothesis tests to allow them to observe unexpected effects. Research questions should be listed under each hypothesis. Primary research questions address high-level effects of the policy. Subsidiary research questions help states and CMS to understand observed effects on outcomes of interest in greater depth and detail.

Relevant hypotheses and research questions are listed in the design appendix for each eligibility and coverage policy. States may also add (1) hypotheses and research questions designed to evaluate unique or state-specific aspects of their policies, or (2) research questions about demonstration processes and implementation that draw on monitoring data or qualitative data and that are not already included in the appendix tables.

C. Methodology

1. Evaluation design summary. In this section of the evaluation plan, the state should provide a brief narrative summary of the proposed evaluation design. Details on the proposed methodology should be described in the following sections and reflected in one or more design tables submitted as part of the evaluation plan. Appendices for individual eligibility and coverage policies contain suggested approaches to address each research question. For research questions included in the state's evaluation design plan, CMS anticipates that states will select at least one approach to each question, although multiple options are presented to prompt state thinking about feasible approaches, given data availability and demonstration design. The use of multiple approaches can serve to strengthen confidence in findings. The state may also propose alternative approaches consistent with best practices in policy evaluation.⁴ The design tables submitted to CMS as part of the state's evaluation plan should include only the approaches the state intends to use after discussing potential approaches with an independent evaluator. Design tables must separately specify comparison strategies, outcome measures, data sources, and analytic approaches for each research question, and must present these components for each research question in the same table so that CMS can assess how well they will work together for a given analysis.⁵

⁴ See "Best Practices in Causal Inference for Evaluations of Section 1115 Eligibility and Coverage Demonstrations" for a general discussion of evaluation planning criteria (<https://www.medicaid.gov/medicaid/section-1115-demo/downloads/evaluation-reports/causal-inference.pdf>) and "Selecting the Best Comparison Group and Evaluation Design: A Guidance Document for State Section 1115 Demonstration Evaluations" for a detailed discussion of appropriate evaluation designs based on comparison group strategies (<https://www.medicaid.gov/medicaid/section-1115-demo/downloads/evaluation-reports/comparison-grp-eval-dsgn.pdf>).

⁵ The format of the design tables recommended for eligibility and coverage demonstrations is adapted from the format in general CMS guidance (<https://www.medicaid.gov/medicaid/section-1115-demo/downloads/evaluation-reports/developing-the-evaluation-design.pdf>).

2. Target and comparison populations. The evaluation plan should provide details on the eligibility groups subject to each eligibility and coverage policy and the inclusion and exclusion criteria for the target group. In addition, for certain eligibility and coverage policies like community engagement, the evaluation plan should discuss how the state will include former beneficiaries in the target group for evaluation purposes. Although the target population for these policies is generally current Medicaid beneficiaries, certain desired outcomes impact both current and former beneficiaries.

Comparison group selection should be informed by the best opportunity to establish a counterfactual to test the effects of the demonstration. This requires identifying a group of individuals who are similar to the demonstration group in their observable characteristics and not subject to eligibility and coverage policies. Ideally, states should identify both within-state and other-state comparison strategies. The state's description of all comparison groups should specify inclusion and exclusion criteria and describe how the state will overcome drawbacks in specific comparison group strategies. For example, eligibility groups not subject to the demonstration could be considered as a comparison group option but may be different from eligibility groups that are part of the demonstration. In some cases, it may be advisable to use propensity score matching or other statistical techniques to reduce dissimilarities in observable characteristics.⁶

If a credible comparison group is difficult to identify for a particular research question, the state should engage with CMS for guidance. CMS will consider, on a case-by-case basis, analytic approaches that do not require comparison groups. For example, information on treatment group outcomes prior to demonstration implementation will allow for interrupted time series analysis, which may be an acceptable approach when credible comparison groups are unavailable. Multiple pre-period observations are necessary for interrupted time series analysis and using the first year of implementation is not sufficient.⁷

Further considerations for in-state comparison groups. Some states may be able to design and conduct a randomized controlled trial (RCT), which randomizes beneficiaries to the demonstration (or to the control group) and therefore creates the best opportunity for unbiased evaluation results. States can also use a “stepped wedge” design, which randomly assigns clusters of beneficiaries to the demonstration at regular intervals, comparing all clusters (those who have and have not received the intervention) at each interval.⁸ States that do not pursue these options should consider staging implementation by geographic area or beneficiary

⁶ In difference-in-differences analyses, similarity of pre-period trends between demonstration and comparison groups is more important than similarity of observable characteristics, and propensity score matching can be counterproductive. For a discussion of bias introduced by matching in difference-in-difference analysis, see Daw, J. R. and Hatfield, L. A. (2018).

⁷ In contrast, having pre-period data on both treatment and comparison group members permits more rigorous methodologies (e.g., difference-in-differences), which are possible with only a single pre-period observation at demonstration baseline.

⁸ These comparison group options are discussed in more detail in separate guidance on designing implementation to support rigorous evaluations. This guidance will be made available at the following link in 2019: <https://www.medicaid.gov/medicaid/section-1115-demo/evaluation-reports/index.html>

characteristics so that beneficiaries not yet exposed to the demonstration can serve as a within-state comparison group for early implementation groups.⁹

Further considerations for other-state comparison groups. States can develop other-state comparison strategies when using national survey data, and should also consider the possibility of drawing on Medicaid administrative data from other states. Comparison states should have a Medicaid program that is similar to the demonstration state in terms of income eligibility guidelines and timing of major eligibility expansions, but should not have a section 1115 demonstration with the same eligibility and coverage policies.

3. Evaluation period. The evaluation plan should clearly specify the evaluation time period, as well as which hypotheses, research questions, and data will be included in both the interim and summative (final) evaluation reports. If the state does not anticipate being able to address certain research questions in the interim evaluation report—either because of a lag in data availability or because outcomes are expected to occur after the interim evaluation period—the evaluation plan should denote those research questions and explain the expected timing.

4. Data sources. In this section, the state should provide details on the data sources for the evaluation, including all data sources listed in the design table(s) that the state submits to CMS. CMS recommends use of both qualitative and quantitative data, depending on the best way to address specific research questions. The state should ensure that proposed data sources can generate planned outcome measures, including, where relevant, for those who were initially in the demonstration population but who have separated from Medicaid. For all measures, the state should include statistical power calculations in the evaluation design plan to show that samples will be of sufficient size to support proposed quantitative analyses. If power calculations suggest that analyses will be underpowered and it is not possible to generate larger samples, the state should engage with CMS for guidance on acceptable use of descriptive analyses. The state should also discuss expected data quality and limitations, the period for which data will be available, and the frequency of collection for each data source. Potential data sources for eligibility and coverage demonstration evaluations include:

- Individual and group interviews with beneficiaries and/or key informants
- Beneficiary surveys, particularly longitudinal surveys that follow current and former beneficiaries over time
- National surveys
- Medicaid administrative data, including claims, encounters, enrollment, and demonstration monitoring data
- Administrative data for non-Medicaid programs, including the Supplemental Nutrition Assistance Program and Temporary Assistance for Needy Families

⁹ Policies that differ on either side of a threshold value for continuous beneficiary characteristics like age or income also create the possibility of conducting regression discontinuity analysis, another methodologically rigorous approach. Regression discontinuity approaches can be combined with implementation staging approaches, but can also be considered in the absence of staged implementation. See “Selecting the Best Comparison Group and Evaluation Design: A Guidance Document for State Section 1115 Demonstration Evaluations” (<https://www.medicaid.gov/medicaid/section-1115-demo/downloads/evaluation-reports/comparison-grp-eval-dsgn.pdf>).

Further considerations for the use of qualitative data. Qualitative data on beneficiary experience and demonstration operations are important for informing quantitative analyses and interpretation of results. States can use qualitative data gathered through individual or group interviews with beneficiaries (sometimes called focus groups) to inform the development of beneficiary survey instruments. For example, if beneficiary interviews yield information on barriers to compliance, states can then develop survey questions to understand the prevalence of those barriers. Interviews with key informants, such as field operations staff and consumer groups, provide information on demonstration operations and implementation experience that can provide context for observed outcomes. States planning to collect qualitative data through individual or group interviews should describe the process for selecting the sample (including the rationale for how participants will be selected, and potentially stratified by characteristics of interest), the targeted numbers of interviewees, and any incentives to be used in recruitment. States should also describe how they plan to use qualitative data in the evaluation, with a focus on how they will integrate qualitative and quantitative data.

Further considerations for beneficiary surveys. Beneficiary surveys developed and fielded by states have different advantages than Medicaid administrative data and national surveys. Beneficiary surveys can yield rich information on beneficiary experiences, barriers to participation in demonstration policies, and understanding of demonstration requirements. They are particularly important data sources for community engagement demonstration evaluations because states must track beneficiaries after they separate from Medicaid to understand employment, income, health status, and coverage transitions over time. Surveys may also be the most accurate source of data on employment and income even if states are able to follow beneficiaries using state workforce or tax data, because some low-income workers do not file taxes, or may have contingent or temporary employment they do not report to states other than to satisfy community engagement requirements. However, if it is possible to gather data on a particular outcome through a reliable non-survey source, states should consider prioritizing the non-survey source due to the need to manage survey length, beneficiary burden, and response rates (which are negatively impacted by survey length).

States that conduct surveys of current and/or former beneficiaries often conduct them as single cross-sectional surveys, which gather data at a point in time. However, understanding certain outcomes of eligibility and coverage policies requires that states administer longitudinal or panel surveys, involving multiple survey waves that follow the same people over time.¹⁰ For example, multiple years of longitudinal data may be necessary to see uptake and maintenance of commercial coverage, which may be conditional on other desired outcomes like employment. These changes may take time to occur. The persistence of these outcomes is also of interest – that is, are coverage changes sustained for at least a year? CMS strongly recommends use of longitudinal beneficiary surveys and is available to provide further technical assistance to states if requested. Many mentions of state beneficiary surveys in the policy-specific appendices assume the survey design is longitudinal.

¹⁰ CMS is developing separate guidance on designing beneficiary surveys to support rigorous evaluations. This guidance will be made available at the following link in 2019: <https://www.medicaid.gov/medicaid/section-1115-demo/evaluation-reports/index.html>

In addition, states should plan to survey beneficiaries at baseline (e.g., before or at demonstration implementation) to support use of difference-in-differences analyses, if they are not able to conduct a randomized controlled trial or regression discontinuity design.^{11,12} Ideally, states should collect baseline data during the period before demonstration implementation. However, it may be acceptable to conduct a baseline survey after demonstration implementation has started, as long as data collection takes place before demonstration policies are expected to affect beneficiaries' behavior or other outcomes. States should also plan to include annual replacement samples of new enrollees so that surveys capture information about the experiences of beneficiaries who enroll after initial implementation.

The description of a planned beneficiary survey in the evaluation plan submitted to CMS must describe the sampling strategy and estimated number of completed surveys, power calculations, stratification by subgroups of particular interest, frequency and timing of data collection, and the method of data collection. The evaluation plan should also describe how the state will meet challenges such as reaching hard-to-reach populations, achieving sufficient response rates, sample weighting for survey nonresponse, and avoiding survey attrition among beneficiaries, especially those disenrolled. Draft survey instruments must be approved by CMS before implementation.

Further considerations for national survey data. Existing national household surveys have several advantages as a data source. They are low-cost, include items relevant to eligibility and coverage policies, allow for comparisons with other states, and provide data for years both before and after demonstration implementation. Drawbacks of national surveys include that they do not allow states to isolate demonstration populations, are known to undercount Medicaid beneficiaries,¹³ and may not have sufficient sample sizes for the population of interest (for example, low-income adults ages 19-64). States should determine whether samples for their population of interest contain enough observations to adequately power planned analyses. In addition, most national surveys do not provide data until the year after they are fielded; depending on the state's evaluation period, it may be more feasible to include analyses of national survey data in summative reports than interim reports. National household surveys listed in the policy-specific appendices include:

¹¹ The planned analytic approach and comparison group strategy may influence the beneficiary survey design. Randomized controlled trials, discussed in section C.5, do not require baseline data because they are an experimental design. Regression discontinuity designs also do not require baseline data, but they do require a sufficient number of observations at the policy eligibility threshold to support analysis; this may be challenging if the data source is a survey.

¹² It is also possible to field a beneficiary survey after implementation and to ask retrospective questions about employment and other beneficiary outcomes before implementation. However, such surveys are subject to recall bias.

¹³ This issue is less impactful where surveys are used to look at the likely eligible population rather than those reporting Medicaid enrollment. It is particularly important to examine the likely eligible population, rather than the population reporting Medicaid enrollment, for evaluations of retroactive eligibility waivers and non-eligibility periods, which may affect outcomes among current beneficiaries as well as the probability of enrollment or the composition of the enrolled population.

- The Integrated Public Use Microdata Sample (IPUMS) version of the American Community Survey (ACS).¹⁴ IPUMS is a research-ready version of the ACS prepared by the Minnesota Population Center at the University of Minnesota. The ACS is administered by the U.S. Census Bureau and has the largest sample size of all federal surveys; samples are large enough to support both state-specific and local-area annual estimates. The ACS provides annual data on employment, education, health insurance, demographic characteristics, and other variables. Data are collected throughout the year using 12 independent monthly samples; estimates may be thought of as an average for the year for a particular geographic area. Data for each calendar year are available in the fall of the following year.
- The Behavioral Risk Factor Surveillance System (BRFSS).¹⁵ BRFSS was established by the Centers for Disease Control and Prevention but is administered by state health departments; it surveys adults ages 18 and older about their health status and is large enough to generate annual state-level estimates. Point-in-time questions are asked on a rolling basis throughout the year. States can choose to use optional survey modules and to add state-specific questions that can inform their evaluation designs, although the lead time to add novel questions can be lengthy since the survey is administered once annually. Data for each calendar year are available in the summer of the following year.
- The Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC). CPS ASEC is administered by the U.S. Census Bureau and collects information on income and other variables using questions about the previous calendar year.¹⁶ The CPS ASEC sample size is smaller than the ACS sample and does not allow for within-state comparisons. Annual state-level estimates are possible but the Census Bureau recommends use of the ACS (not the CPS) for one-year estimates of income by demographic characteristics.¹⁷ If needed variables are available in the ACS, the ACS should be used instead. CPS data for each calendar year are available in the fall of the following year.

States may also propose to use national surveys not listed in the policy-specific appendices, such as the National Health Interview Survey,¹⁸ if samples are large enough for their states and population of interest. In addition, states may wish to use national surveys, including those with small sample sizes, as a source of tested survey questions for state-based beneficiary surveys.¹⁹

Further considerations for Medicaid administrative data. Medicaid administrative data, including claims, encounters, and enrollment data, are necessary to answer research questions about the effects of eligibility and coverage policies on utilization and enrollment. Medicaid

¹⁴ IPUMS-ACS: <https://usa.ipums.org/usa/acs.shtml>

¹⁵ BRFSS: <https://www.cdc.gov/brfss/about/index.htm>

¹⁶ CPS ASEC: <https://www.census.gov/programs-surveys/saipe/guidance/model-input-data/cpsasec.html>

¹⁷ See <https://www.census.gov/topics/income-poverty/poverty/guidance/data-sources.html>

¹⁸ The National Health Interview Survey is administered by the National Center for Health Statistics. See https://www.cdc.gov/nchs/nhis/about_nhis.htm

¹⁹ Forthcoming guidance on state-based beneficiary surveys will include a bank of survey items from existing national and state surveys. This guidance will be made available at the following link in 2019: <https://www.medicare.gov/medicaid/section-1115-demo/evaluation-reports/index.html>

administrative data can support rigorous quasi-experimental evaluation designs because large numbers of observations are typically available both before and after demonstration implementation.²⁰ States that rely on comparison groups of beneficiaries in other states may need to request Medicaid administrative data from other states. Medicaid administrative data may also include data on demonstration operations, including monitoring data, that are helpful for answering research questions about demonstration processes. Note that monitoring data alone are not sufficient for evaluation purposes because states are expected to apply more robust econometric methods to address research questions about expected demonstration outcomes, such as using comparison groups and statistical techniques to control for observable beneficiary characteristics.

Further considerations for non-Medicaid administrative data. States should assess whether non-Medicaid sources of administrative data can generate planned outcome measures and be linked to Medicaid data. This is particularly important for states with community engagement requirements due to the need to follow beneficiaries after separation from Medicaid. For example, states should plan to access data systems for public programs such as Temporary Assistance for Needy Families (TANF) and the Supplemental Nutrition Assistance Program (SNAP) to understand whether community engagement requirements affect participation in these programs. States may wish to link Medicaid to state workforce or tax data to understand employment and income outcomes; however, such data may exclude individuals who meet work requirements but do not file income taxes or otherwise report their income. State unemployment insurance filings are another potential source of employment information. All-payer claims database (APCD) data may serve as a source of information on enrollment commercial health insurance, although relatively few states have APCDs, and APCDs vary in (1) whether they use a unique person identifier that can track transitions across insurance coverage types, (2) the quality of the unique person identifier, and (3) how many insurance carriers report to the APCD. In general, states planning to draw on administrative data for non-Medicaid programs should facilitate evaluators' access to these data by coordinating with relevant state agencies.

5. Analytic methods. In this section, the state should provide details on all analytic methods planned for the evaluation, including each analysis listed in the design table(s) as submitted to CMS. CMS suggests a mixed-methods evaluation design, using both qualitative and quantitative data to conduct descriptive and impact analyses. The objective of qualitative analyses is to understand demonstration operations and beneficiary experience, and to support the design and interpretation of quantitative descriptive and impact analyses. The objective of quantitative analyses is to assess measured changes in demonstration outcomes, some of which may be attributable to the demonstration. The state should explain how the evaluation will integrate findings from both types.

Further considerations for quantitative analyses. Suggested quantitative analyses in the policy-specific appendices include experimental, quasi-experimental, and descriptive

²⁰ Medicaid administrative data for the period before demonstration implementation may not be available for demonstrations that coincide with eligibility expansions, since it is not feasible to collect retrospective data on beneficiary outcomes for individuals newly eligible for Medicaid. In these cases, difference-in-differences analysis is not possible; states may therefore want to use regression discontinuity designs or incorporate randomized controlled trials.

approaches. States can consult “Best Practices in Causal Inference for Evaluations of Section 1115 Eligibility and Coverage Demonstrations” and “Selecting the Best Comparison Group and Evaluation Design: A Guidance Document for State Section 1115 Demonstration Evaluations”²¹ for detailed discussions of analytic strategies, how they depend on available comparison groups, and the extent to which they may support causal attribution.

- Randomized controlled trials (RCTs) are experimental and allow causal attribution of observed effects to demonstration policies if executed correctly. Because assignment is random, the only difference between the treatment group (i.e., the demonstration group) and the control group is the exposure to the demonstration. Therefore, estimates of demonstration effects will not be biased by characteristics of beneficiaries selected into the treatment group. CMS recommends use of RCTs, but states should consider both the benefits to this approach and the planning and investment required for correct execution and application to specific research questions. For example, using an RCT to analyze subgroup effects will require stratified randomization so that there is balance on beneficiary characteristics across subgroups. Conducting an RCT minimizes, though may not eliminate, the need to collect baseline data.²²
- Quasi-experimental approaches such as regression discontinuity designs and difference-in-differences analysis may support causal inference, depending on the strengths and limitations of the specific analysis. States may choose these approaches based on data and comparison group availability, but should also consider the tradeoffs between them; for example, regression discontinuity designs offer strong internal validity but results do not necessarily generalize to the entire population subject to a demonstration. States should consider staggering implementation by geography or beneficiary characteristics such as age to support stronger quasi-experimental evaluation design approaches.
- Descriptive analysis is an appropriate method for addressing research questions about demonstration *processes*. Descriptive analyses of demonstration *outcomes* may be necessary if no comparison group is available to address a particular research question, if power calculations suggest that a sample will be too small to observe a statistically significant effect, or if a state is making multiple comparisons with the same data source, which can lead evaluators to incorrectly attribute observed outcomes to demonstration policies. Descriptive analyses, including descriptive regressions, do not support causal inference about the effects of demonstration policies and states must interpret results with caution. CMS will consider, on a case-by-case basis, analytic approaches that are not experimental or quasi-experimental and do not require comparison groups.

²¹ “Best Practices in Causal Inference for Evaluations of Section 1115 Eligibility and Coverage Demonstrations” is available at <https://www.medicaid.gov/medicaid/section-1115-demo/downloads/evaluation-reports/causal-inference.pdf> and “Selecting the Best Comparison Group and Evaluation Design: A Guidance Document for State Section 1115 Demonstration Evaluations” is available at <https://www.medicaid.gov/medicaid/section-1115-demo/downloads/evaluation-reports/comparison-grp-eval-dsgn.pdf>.

²² The expectation is that random assignment ensures that demonstration and control groups are similar; however, there is some random chance they will not be—particularly if sample sizes are small. Baseline data can be used to provide further statistical adjustment to ensure the two groups are similar. States may also wish to conduct further quasi-experimental analyses using the same control group created through randomization for experimental analyses, in which case baseline values are needed.

Further requirements for the evaluation plan. For each planned analytic approach included in the state’s design table(s) as submitted to CMS, the evaluation plan should describe the target population, time points for data collection and outcome measures. For planned regressions, the evaluation plan should specify the statistical model and list control variables. Depending on the policy being evaluated and the expected outcome, important control variables may include demographics, underlying health status, access to health care services, availability of jobs, and job characteristics.

Subgroup analyses. Suggested research questions in each policy-specific design appendix do not include questions about subgroup effects. States should work with CMS to define important subgroups for their demonstration design and context. For example, it may be important to examine whether observed effects on an outcome like employment differ for those enrolled in Medicaid for different lengths of time, by regional economic conditions, by substance use disorder status, by disability status, by racial/ethnic group, by age group, and/or by eligibility group. States should plan to gather enough data to allow observation of differences in relevant subgroup outcomes of at least 5 percent within 95 percent confidence intervals.

D. Methodological Limitations

In this section, states should provide detailed information on the limitations of the evaluation. These could include the availability of comparison groups, the data sources or collection process, or analytic methods. The state should also identify any efforts to minimize limitations, and should acknowledge where limitations will preclude causal inferences about the effects of demonstration policies. In addition, this section should include any information about features of the demonstration that present methodological constraints that the state would like CMS to consider in its review.

E. Attachments

1. Independent evaluator. Describe the process the state will use to work with its contracted independent evaluator to conduct the analysis and write the evaluation report. Explain how the state has involved the independent evaluator in the development of the evaluation plan. Explain how the state will assure that the independent evaluator will conduct a fair and impartial evaluation, prepare an objective evaluation report, and that there would be no conflict of interest. The evaluation design should include “No Conflict of Interest” signed confirmation statements from the independent evaluator.

2. Evaluation budget. A budget for the evaluation shall be provided with the draft evaluation design. It will include the total estimated cost, as well as a breakdown of estimated staff, administrative, and other costs for all aspects of the evaluation. Examples include, but are not limited to: the development and fielding of all survey and measurement instruments, quantitative and qualitative data collection, data cleaning and analyses, and reports generation. A justification of the costs may be required by CMS if the estimates provided do not appear to sufficiently cover the costs of the draft evaluation design or if CMS finds that the draft evaluation design is not sufficiently developed.

3. Timeline and major milestones. Describe the timeline for conducting the various evaluation activities, including dates for evaluation-related milestones, including those related to

procurement of an outside contractor, if applicable, and deliverables. The final evaluation design shall incorporate an interim and summative evaluation report. Pursuant to 42 CFR 431.424(c)(v), this timeline should also include the due date for the final summative evaluation report.

References

- Colby, M., K. Bradley, K. Contreary, and B. Natzke. “Medicaid 1115 Demonstration Evaluation Design Plan: Premium Assistance Monthly Payment, and Beneficiary Engagement. Design Supplement: Interim Outcomes Evaluation.” Report submitted to the Centers for Medicare & Medicaid Services, July 2017. Available at <https://www.medicaid.gov/medicaid/section-1115-demo/downloads/evaluation-reports/eval-plan-beneficiary-engagement-programs.pdf>
- Contreary, K., K. Bradley, and S. Chao. “Best Practices in Causal Inference for Evaluations of Section 1115 Eligibility and Coverage Demonstrations.” Report submitted to the Centers for Medicare & Medicaid Services, June 2018. Available at <https://www.medicaid.gov/medicaid/section-1115-demo/downloads/evaluation-reports/causal-inference.pdf>
- Daw, J. R. and Hatfield, L. A. (2018) “Matching and Regression to the Mean in Difference-in-Differences Analysis.” *Health Services Research*, 53: 4138-4156. doi:[10.1111/1475-6773.12993](https://doi.org/10.1111/1475-6773.12993)
- Reschovsky, J.D., J. Heeringa, and M. Colby. “Selecting the Best Comparison Group and Evaluation Design: A Guidance Document for State Section 1115 Demonstration Evaluations.” Report submitted to the Centers for Medicare & Medicaid Services, June 2018. Available at <https://www.medicaid.gov/medicaid/section-1115-demo/downloads/evaluation-reports/comparison-grp-eval-dsgn.pdf>