CMS
CENTERS FOR MEDICARE & MEDICAID SERVICES

Evaluation Design Resources for Section 1115 Medicaid Demonstrations

Learning Collaborative

July 31, 2019
1:30 PM – 3:00 PM ET

Danielle Daly, CMS
Paula Kazi, CMS
Kate Bradley, Mathematica

Agenda

I. Monitoring and Evaluation of Section 1115 Demonstrations

II. Overview of Technical Resources for Demonstration Evaluations

III. Deep Dive: Implementation to Support Evaluation Design

Question & Answer

IV. Deep Dive: Beneficiary Survey Design Guidance

Question & Answer

Integrated approach to monitoring and evaluation

- All section 1115 demonstrations are policy experiments that must be carefully implemented, monitored, and evaluated (CFR 432.420, 432.424, 432.428)
- With robust and more standardized reporting, states can make course corrections as needed, and learning within and across states can be accelerated
- CMS has developed monitoring and evaluation resources for community engagement and other eligibility and coverage policies to provide clear guidance to states
  - Resources for demonstrations addressing Substance Use Disorder (SUD) are also available
  - Resources for demonstrations addressing Serious Mental Illness (SMI) and Serious Emotional Disturbance (SED) are in progress

Monitoring and evaluation tools for states

- Templates are specific to demonstration policies – content varies by policy type, and not all policies have each template
- CMS has observed uptake of demonstration policy-specific evaluation design guidance released this spring (e.g., SUD, community engagement)

Template	Description
Implementation Plan	▪ Documents state approach to implementation ▪ Informs monitoring and evaluation activities for the demonstration
Monitoring Protocol	▪ Describes what state will report on a quarterly basis, developed collaboratively between CMS and the state
Monitoring Report	▪ Documents qualitative summaries on metrics trends and implementation updates ▪ Provides standardized quarterly and annual monitoring metrics
Evaluation Design Guidance	▪ Supports states in developing evaluation designs that meet CMS expectations for rigor ▪ Recommends hypotheses, research questions, outcome measures, and analytic approaches

Overview of technical resources for section 1115 demonstration evaluations

Resources providing an introduction to evaluation design possibilities and principles

Paper #1: <u>Selecting the Best Comparison Group and Evaluation Design</u>

- Overview of comparison group options, focusing on ***quasi-experimental*** evaluations
- Includes high-level discussion of statistical approaches necessary to ensure appropriate inferences when using comparison groups
- Discusses ***non-experimental designs*** when no comparison groups are available

Quasi-experimental designs are observational studies that identify a comparison group that
1. Is not subject to the demonstration policy, and
2. Has similar baseline characteristics compared to the group subject to the policy

Non-experimental designs typically lack either a comparison group or baseline observations. Examples of non-experimental designs include pre-test/post-test, post-test only, or post-test with non-equivalent groups used for comparison.

Resources providing an introduction to evaluation design possibilities and principles *(continued)*

Paper #2: <u>Best Practices in Causal Inference</u>

- In section 1115 demonstrations, causal inference is the process of determining whether a demonstration policy is responsible for an observed outcome
- This paper discusses how to:
  - Establish valid counterfactuals (i.e., what would have happened in the absence of the demonstration)
  - Check the feasibility of causal inference
  - Use multiple methods to interpret results
  - Check the robustness of findings (i.e., whether results should be trusted and how generalizable they are)
- Includes examples relevant to eligibility and coverage demonstrations, common pitfalls, and practical suggestions

Resources on advanced evaluation design topics

Paper #3: <u>Planning Section 1115 Demonstration Implementation to Enable Strong Evaluation Designs</u>

- Describes how states can plan ahead so implementation can support:
  - Experimental designs (such as randomized controlled trials, or RCTs)
  - Comparison group options for quasi-experimental designs
  - Baseline data collection
  - Methods to disentangle the effects of specific demonstration features

Resources on advanced evaluation design topics

Paper #4: <u>Beneficiary Survey Design and Administration for Eligibility and Coverage Demonstration Evaluations</u>

- Describes considerations and suggested approaches for designing beneficiary surveys, including:
  - Developing a sampling plan
  - Selecting survey mode
  - Developing instruments
  - Fielding surveys
- Appendices include suggested survey items that correspond to measures recommended in CMS evaluation guidance for section 1115 eligibility and coverage policies

Audience Poll!

Which of the four resources are you most interested in to support your state's section 1115 demonstration evaluation?

A. Selecting the Best Comparison Group and Evaluation Design

B. Best Practices in Causal Inference

C. Planning Section 1115 Demonstration Implementation to Enable Strong Evaluation Designs

D. Beneficiary Survey Design and Administration for Eligibility and Coverage Demonstration Evaluations

Deep dive: Implementation to support evaluation design

Four key advantages to planning implementation with evaluation in mind

- Enable experimental evaluation
- Support baseline data collection
- Expand options for quasi-experimental evaluation
- Disentangle effects of different demonstration policies

Enable experimental evaluation

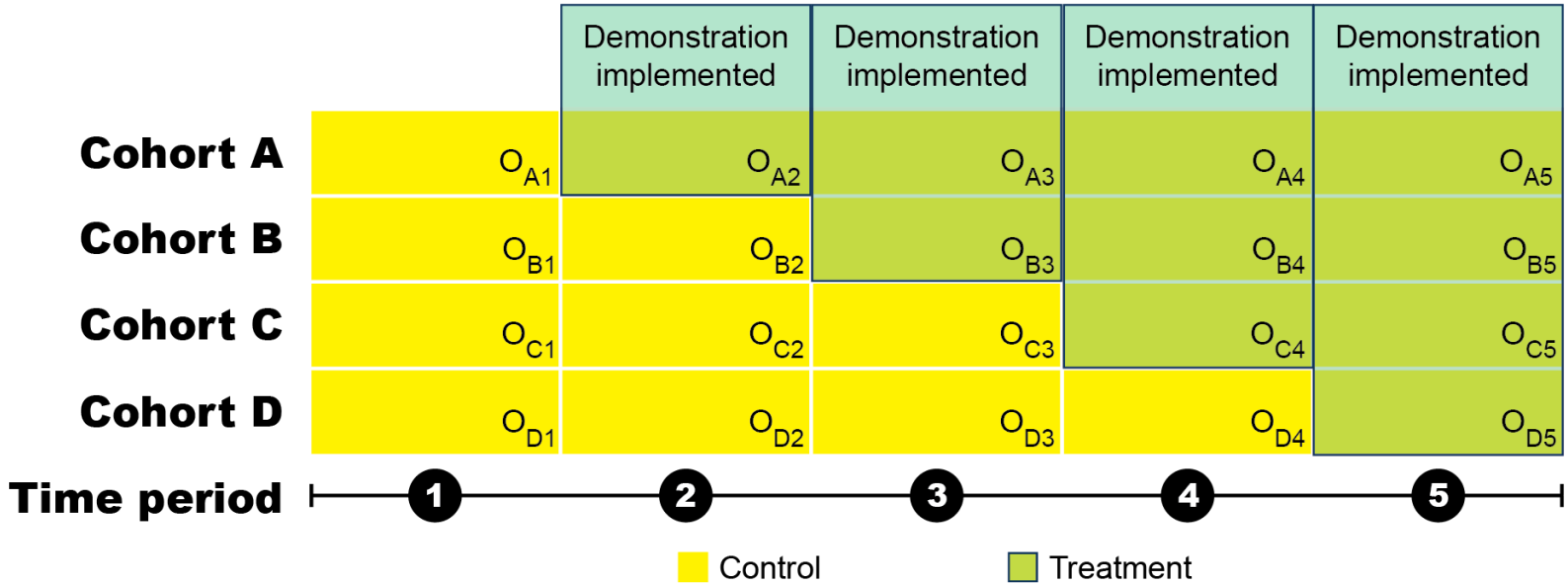
- States interested in randomized controlled trials (RCTs) will need to plan early and conduct random assignment before implementation
  - Must also decide how to assign beneficiaries who become eligible for the demonstration after it has started
- Advantageous for demonstrations with eligibility and coverage policies since beneficiary response is likely to be influenced by unobserved factors
  - Community engagement requirements promote work, but inherent motivation to work and relevant labor market skills vary among beneficiaries
- Administration of RCTs may be costly, but potentially offset by more straightforward data analysis and no requirement for baseline data as with quasi-experimental approaches

Support baseline data collection

- Baseline data support rigorous quasi-experimental evaluation designs like difference-in-differences
- Administrative data may be available before implementation, but many demonstrations test outcomes that cannot be measured with Medicaid administrative data
  - For example, transitions to commercial insurance, employment outcomes, and long-term health status
- Because individual-level data are important to isolate demonstration effects, states advised of opportunity to collect data from beneficiaries at demonstration baseline (likely by survey), before implementation
  - Baseline period could also be after implementation, but before policy might affect beneficiary behavior
  - For example, the months between initial implementation and distribution of first premium invoices, which may be first awareness of costs

Expand options for quasi-experimental evaluation: stepped wedge design

- Stepped wedge design allows beneficiaries in later implementation cohorts to serve as a comparison group for beneficiaries selected for earlier implementation



Note: Observations (O) on outcomes and control variables are made for each cohort five times, once at the end of each time period.

Expand options for quasi-experimental evaluation: regression discontinuity design

- Beneficiaries subject to eligibility and coverage policies typically defined by demographic characteristics, such as age or income
- Regression discontinuity designs provide estimates of impacts on beneficiaries who are close to the eligibility threshold value. State could phase in implementation to create multiple discontinuities. For example:
  - Policy applies to all working age adults ages 20-64
  - Implement three age groups sequentially to create two discontinuity thresholds
    - 20-34, 35-49, 50-64
    - Thresholds at age 35 and 50

Disentangle effects of different demonstration policies

- Factorial design randomizes beneficiaries to all possible policy combinations
  - Allows assessment of each policy and interactive effects
- What is the effect of community engagement requirements?
  - Compare groups 5-8 vs. groups 1-4
- Does impact of premiums vary if non-eligibility periods are also in place?
  - Compare groups 3 and 7 vs. groups 4 and 8

	Community engagement		Premiums		Non-eligibility periods	
Evaluation group	Yes	No	Yes	No	Yes	No
1		X		X		X
2		X		X	X	
3		X	X			X
4		X	X		X	
5	X			X		X
6	X			X	X	
7	X		X			X
8	X		X		X	

Questions?

Deep dive: Beneficiary survey design guidance

Audience Poll!

At what stage are you in planning a beneficiary survey for your evaluation?

A. Do not intend to survey beneficiaries

B. Planning to conduct a beneficiary survey, but have not started designing

C. In the planning and design process for a beneficiary survey

D. Beneficiary survey plans are complete, but not yet fielded

E. Beneficiary survey is already underway

Importance of beneficiary surveys

- Surveys can yield rich information about beneficiaries' understanding of and experiences with eligibility and coverage policies
  - Inform interpretation of other evaluation results and could help improve implementation of the demonstration
- Help states assess changes in beneficiary outcomes over time, including for those who disenroll or transition to other coverage
  - Longitudinal outcomes through repeated observations shed light on how long it takes for desired outcomes to occur and the persistence of those outcomes
  - In states without access to all-payer claims databases, plausibly offers the best source of data-for assessing long-term outcomes
- Surveys are the recommended data source for many research questions in CMS' policy-specific evaluation design guidance

Audience Poll!

For which of the following eligibility and coverage demonstration evaluation topics is a beneficiary survey the most appropriate data source?

A. Changes in continuity of beneficiary enrollment

B. Beneficiary understanding of waiver policies

C. Changes in utilization patterns of inefficient health care services

D. Not sure

Topics covered in beneficiary survey design guidance paper: survey design

- Survey design options
  - Type of survey design that will generate needed data
  - When and how many times the survey should be fielded
  - Length of the field period
- Creating a sampling plan
  - Designing a representative sample
  - Sample size and target response rate
  - Survey eligibility criteria
  - Adjustment for nonresponse
- Importance of mode in collecting survey data
  - Key considerations for choosing mode
  - Mixed-mode designs to achieve target response rates
  - Mode choice and data quality considerations

Topics covered in beneficiary survey design guidance paper: developing the instrument

- Moving from research questions to survey questionnaires
  - One vs. multiple survey items for a topic
  - Strategies to develop new items
  - Best practices for item wording
  - Modifying items for different modes
- Preparing instrument to ensure high quality data
  - Handling completed and qualified partial cases
  - Surveying non-English speakers
  - Allowing proxy interviews
  - Computer Assisted Telephone Interviewing (CATI) formats and testing the programmed instrument
- Testing survey instruments and procedures
  - Pre-testing and pilot testing

Topics covered in beneficiary survey design guidance paper: fielding the survey

- Sample management
  - Ensures appropriate contact, routing, and follow-up with each sample member, and accurate record-keeping
- Survey field periods with more than one release of sample
- Strategies to achieve target response rates and minimize attrition
- Monitoring the data collection process for quality

Topics covered in beneficiary survey design guidance paper: sample items

- Appendices contain example item wording for measures listed in CMS evaluation design guidance for eligibility and coverage policies
- Draws on federal surveys and publicly available state surveys, with direct links to survey questionnaires

Table C.1. Example survey items that collect demographic data

Measure	Example item(s), [source]	Notes
Age	• What is your age? (Specify age in years) [AGE, BRFSS 2018] • What is your age? (18 to 24; 25 to 34; 35 to 44; 45 to 54; 55 to 64; 65 to 74; 75 or older) [33, CAHPS Adult Medicaid]	A continuous response format may be preferable for analysis, and the data can always be converted to categories. However, the continuous response format may have higher non-response than a categorical response format, especially if the survey is self-administered.
Sex / gender	• Format 1: What is your sex? Format 2: What was your sex at birth? Is it… (Male; Female; Don't know/Not sure) [SEX1, BRFSS 2018] • Which of the following best represents how you think of yourself? (Lesbian or Gay; Straight, that is, not gay; Bisexual; Something else; I don't know the answer; Refused) [SOFEMALE, BRFSS 2018] • Do you consider yourself to be transgender? (Yes – Transgender, male-to-female; Yes – Transgender, female-to-male; Yes – Transgender, gender nonconforming; No) [TRNSGNDR, BRFSS 2018]	BRFSS offers two formats of questions on sex. The BRFSS items on sexual orientation and transgender identification can help identify subgroups who may experience different outcomes.
Primary spoken language	• What language do you mainly speak at home? (English; Spanish; Chinese; Russian; Vietnamese; Portuguese; Some other language – Specify) [32, CAHPS Hospital Survey]	States may want to modify response options to include other high-prevalence languages in their region
Work-disability status	• Because of a physical, mental, or emotional condition lasting 6 months or more, does <this person> have any difficulty working at a job or business? (Yes; No) [DISABWRK, IPUMS ACS 2007]	

Topics covered in beneficiary survey design guidance paper: sample items

Table C.2. Example survey items for evaluations of community engagement requirements

Research question	Measure needed for research question	Example item(s), [source]	Notes
RQs 1.1, 1.1a, and 1.1b	Probability of being employed	Are you currently…? (Employed for wages; Self-employed; Out of work for 1 year or more; Out of work for less than 1 year; a Homemaker; a Student; Retired; Unable to work) [EMPLOY1, BRFSS 2018] Alternatively: A. Last week, did <this person> work for pay at a job (or business)? (Yes; No – Did not work or retired) B. Last week, did <this person> do ANY work for pay, even for as little as one hour? (Yes; No) [EMPSTAT, IPUMS ACS 2017]	BRFSS structures this item as a single response option. Respondents who feel they fit into more than one category are directed to select the category that best describes them. Using the BRFSS item wording and structure allows states to compare their survey data with trends in their state overall. However, responses would not capture the number of people who report being a student and also being employed. If this detail is important, states should modify this question to a "select all that apply" format or consider asking a separate item to ascertain whether a respondent is currently in school.
	Number of hours worked per week	• How many hours did <this person> work LAST WEEK at all jobs? (Specify hours by subtracting any time off and adding overtime or extra hours worked) [HRSWORK1, IPUMS ACS 1990] Alternatively, repeat the following items for all named jobs: • Average number of hours worked per week at job 1 during the reference month. (Specify number of hours) [TJB1_MWKHRS, SIPP 2014] • Average number of hours worked per week at job 2 during the reference month. (Specify number of hours) [TJB2_MWKHRS, SIPP 2014]	These items should be asked only of respondents who report being employed. States should carefully weigh the time period of reference for items asking about job(s). Asking about jobs worked in the prior week poses less recall burden, though it may not capture employment experiences for respondents with sporadic or seasonal employment. States seeking a comprehensive view of current employment should collect data for multiple jobs separately. This can be done by asking whether respondent worked more than one job during time period of reference (e.g., last week) and then asking for detail on each. If not looping through all current jobs, simplify the task by asking respondents to report on the job where they work most hours. This approach also streamlines complexity for paper questionnaires. Asking respondents to sum hours across all jobs, regardless of time period of reference, can be burdensome and lead to inaccurate reporting if they work multiple jobs and/ or schedules that vary from week to week.

Tricky issue #1: Designing adequately powered sample

- Oversampling may be necessary to ensure adequate representation of important subgroups
  - Smaller demographic groups, such as racial/ethnic minorities
  - Particular groups impacted by demonstration policies, such as:
    - Those disenrolled for non-compliance with demonstration policies
    - Higher income groups required to make premium payments
- Power calculations (which will tell you how likely it is that a survey will detect a policy effect that is present) are necessary to support statistically sound analyses
  - Underpowered analyses may fail to detect real policy effects
  - Overpowered analyses can waste evaluation resources
  - Appendix B describes information needed for power calculations, but recommends states rely on their evaluators for the statistical expertise

Tricky issue #2: Using multiple survey modes

- Barriers to participation for low-income people exist for all survey modes

Mode	Potential barriers to participation
Self-administered web	• Limited digital literacy – trouble logging into or navigating a survey website • Limited literacy • Lack of access to a secure, private place to complete the questionnaire • Concerns about phishing scams • Concerns about data privacy
Self-administered paper	• Transiency – moving often, sometimes without a forwarding address • Housing instability – living with friends or family, impossible to locate using traditional online resources • Limited literacy • Cognitive barriers to navigating skip patterns or following survey instructions • Impaired vision, making it difficult to read the text
Interviewer-administered telephone	• Intermittent phone service • No access to a telephone • Use of pre-paid cell phones, causing person's phone number to change frequently • Calling plans with limited number of minutes per month • Concerns about telephone scams; confusing surveys with telemarketing • Hearing impairments that preclude participation by telephone without use of assistive technologies
Interviewer-administered in-person	• Lack of access to a private place to answer the questions • Transiency – moving often, sometimes without a forwarding address • Housing instability – living with friends or family and cannot be located using traditional online resources • Concerns about personal safety (opening the door to a stranger), scams, confusion with for-profit or other solicitations

Tricky issue #2: Using multiple survey modes

- Using multiple modes sequentially can help overcome some of these barriers and conserve survey resources by allowing evaluators to see how effective each mode is, and for which types of respondents
  - But – different modes introduce complexity in data analysis!
- Each mode has unique data quality concerns that evaluators can proactively address (see Table II.5)
- Evaluators should:
  - Ensure that questionnaires are designed with equivalence across modes
  - Estimate mode effects
  - Adjust for unintended differential mode effects

Tricky issue #3: Ensuring survey data quality

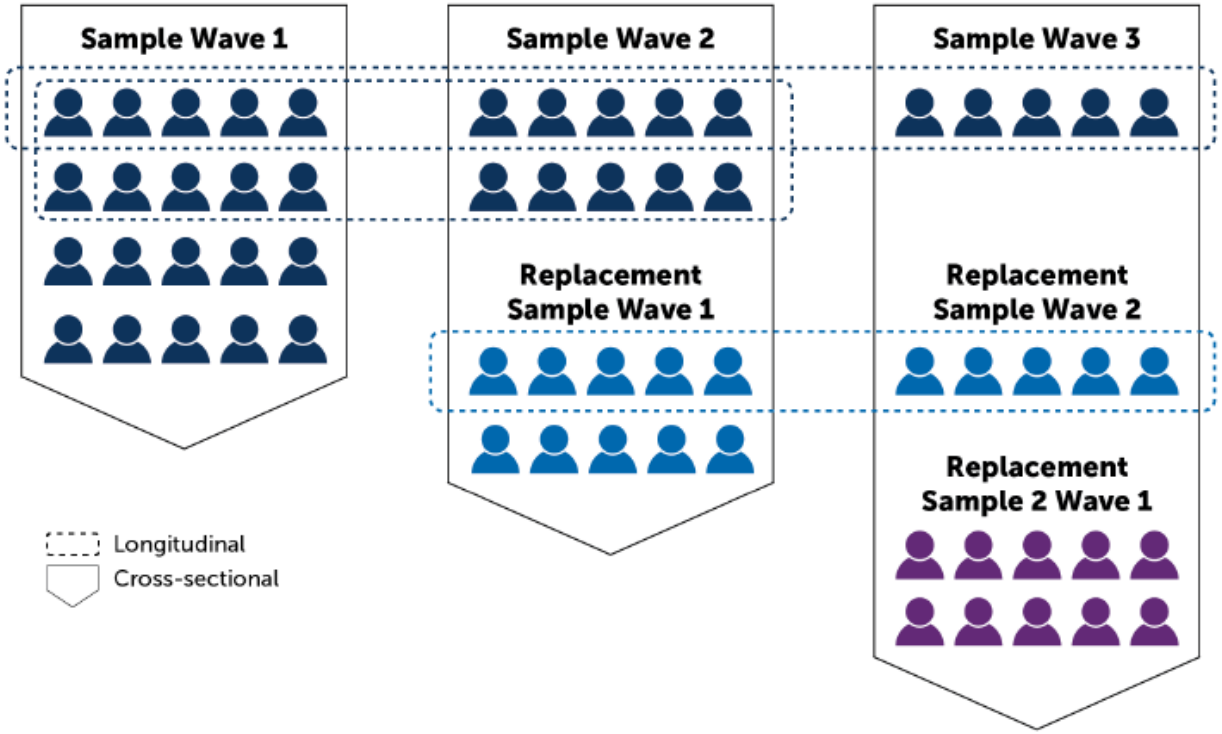
- Use computer-assisted interviews to integrate data quality checks and prevent collection of poor quality data, especially for items considered essential to the planned analyses
  - Allowable ranges for responses
  - Hard checks on responses – unacceptable response or non-response prevents progression through the survey
  - Soft checks on responses – generates alert, but allows progression through the survey
- Analyze data to understand whether responses differ by mode and level of assistance (if proxies are allowed)

Tricky issue #4A: Keeping survey participants engaged

- Low response rates signal higher likelihood of bias – respondents and non-respondents may differ in meaningful ways
- Strategies for keeping participants engaged include:
  - Answer sample members' questions or concerns during the survey
  - Partner with community-based organizations that serve Medicaid beneficiaries
  - Consistent survey branding
  - Advance notification mailings
  - Prepaid or post-paid incentives for participation
  - Non-response follow-up
  - Interim contacts
  - Special efforts to locate hard-to-reach sample members

Tricky issue #4B: Handling attrition and non-response

- Construct and apply survey weights to achieve unbiased estimates
- Consider drawing replacement samples of new enrollees
  - Also allows learning about beneficiaries who enroll after the first survey wave



Audience Poll!

Which of the "tricky issues" in beneficiary survey design are you most concerned about in planning your demonstration evaluation?

A. Designing adequately powered sample
B. Using multiple survey modes
C. Ensuring survey data quality
D. Getting sufficient response rates
E. Handling attrition over time

Questions?

Coming Soon

Upcoming Webinar:

- State experiences with developing Implementation Plans in late August

Final Remarks

- Thank you again for participating in the section 1115 demonstration evaluation design resources webinar!
- We value your time and questions, and encourage you to email for technical assistance: 1115MonitoringAndEvaluation@cms.hhs.gov