May 2023

Federal Meta-Analysis Support: Section 1115 Serious Mental Illness/Serious Emotional Disturbance Demonstrations

Evaluation Design Report

Prepared for
Michael Trieger
Centers for Medicare & Medicaid Services
7500 Security Boulevard, Mail Stop B3-03-30
Baltimore, MD 21244-1850

Prepared by

RTI International
3040 E. Cornwallis Road
Research Triangle Park, NC 27709

RTI Project Number 0214448.001.012

RTI INTERNATIONAL

CONTENTS

LIST OF EXHIBITS

EXECUTIVE SUMMARY

As described in federal statute, Medicaid's primary objective is to provide medical assistance, and the Centers for Medicare & Medicaid Services (CMS) partners with states to provide that medical assistance, with a goal of improving the health of the low-income individuals Medicaid serves through delivery of high quality, equitable care. Section 1115 Medicaid demonstrations offer Medicaid programs an opportunity to design, implement, and test new approaches to their programs that have the potential to further this goal, as well as to improve Medicaid program effectiveness, increase access to health care, and reduce disparities in health care and health so that Medicaid enrollees can achieve the highest level of health. These demonstrations can also shape new policy directions at the federal level.

All section 1115 demonstrations are policy experiments that must be carefully implemented, monitored, and evaluated. To learn from groups of Medicaid section 1115 demonstrations with similar features, the CMS commissioned the Federal Meta-Analysis Support contract. Under this contract, RTI International is working with CMS to conduct meta-evaluations of Medicaid section 1115 substance use disorder demonstrations and serious mental illness/serious emotional disturbance (SMI/SED) demonstrations. This Evaluation Design Report describes the planned meta-evaluation for the SMI/SED demonstration. The meta-evaluation will compare experiences among SMI/SED demonstrations across states to understand the overall effectiveness of the demonstrations and how variation in state demonstration features and the context in which they are implemented contribute to differences in effectiveness. The meta-analyses will primarily use data from state demonstration monitoring and evaluation reports, augmented with limited stakeholder interviews, Medicaid claims data from the Transformed Medicaid Statistical Information System (T-MSIS), and other national datasets.

Through this work, RTI will collaborate with CMS and its other contractors to highlight best practices in implementing demonstration activities and to identify the impacts of those activities to inform national policy making and to support scaling up and diffusion of successful activities and policies. An additional goal of this project is to inform CMS on the rigor and limitations of state evaluations to support further improvements in CMS evaluation guidance for section 1115 demonstrations. By combining an in-depth look at the section 1115 SMI/SED demonstrations' context, implementation, and outcomes, the meta-analysis will complement state evaluations.

ES.1 Section 1115 Demonstrations

The Medicaid program is jointly administered by states and the federal government to provide medical assistance to certain groups, such as low-income individuals and those with certain medical conditions, with a goal of improving health and promoting high quality care. Medicaid is a complex program, serving many disparate groups with very different health needs. Identifying innovative approaches to address high-priority challenges that emerge in administering high quality medical assistance to these groups is of critical interest to state Medicaid programs and their federal partner, CMS. Since the inception of Medicaid, states have been able to use research and demonstrations authorized under section 1115(a) of the Social Security Act to waive certain Medicaid requirements to test changes in how health care services are delivered, which services are covered, which populations are eligible for the Medicaid program, and how providers are paid, all while maintaining the overall goals of the Medicaid program and budget neutrality within the program. CMS has announced its commitment to supporting state innovations in their Medicaid programs and allowing flexibility for states to adapt demonstration design to reflect the uniqueness of their covered populations, resources, and policy goals.

This Evaluation Design Report for the Federal Meta-Analysis Support contract focuses on section 1115 SMI/SED demonstrations. These demonstrations allow states to test novel approaches to delivering SMI services for adults and SED services for youth in Medicaid with a goal of reducing length of stay in emergency departments; reducing readmissions to acute care hospitals and residential settings; improving availability of crisis stabilization, intensive outpatient, psychiatric hospital, and residential treatment setting services; improving access to community-based services and integration of primary and behavioral health care; and improving care coordination and continuity of care after a hospitalization or residential treatment stay.

ES.2 Overview of Meta-Analytic Approach

States are required to monitor and evaluate their section 1115 demonstrations per 42 CFR §431.424 and §431.428. Monitoring provides early and ongoing information about demonstration implementation and progress toward milestones that CMS and states can use to identify potential problems and to make midcourse adjustments if needed. Evaluations seek to understand demonstration implementation progress and impacts of demonstration activities on health care use, health care costs, and beneficiary health. CMS and states can use evaluation findings to make program changes and, as needed, to achieve desired outcomes. However, state-specific monitoring and evaluation cannot identify overarching lessons **across states** that can be used to shape broader state and federal Medicaid policies to improve care for individuals with

SMI/SED. Meta-analyses identify patterns in implementation practices and demonstration impact across states and apply methods to draw out and explain which policies, program components, and contextual factors (collectively referred to as demonstration features) explain more or less success in meeting demonstration goals. Key activities in the meta-analysis of the section 1115 SMI/SED demonstration include:

- Conducting an environmental scan of state program documents and websites to identify demonstration features that can be used in the meta-analysis.
- Conducting interviews with Medicaid agency and behavioral health agency staff to understand a state's demonstration planning process, demonstration features, the changes made in implementation and the facilitators and barriers to making changes, and potential impacts of other initiatives in the state on the demonstration.
- Analyzing state-reported monitoring metrics to assess how states are doing in implementing the demonstration and meeting demonstration milestones, which will complement the information abstracted from demonstration program documents and interviews with state staff.
- Employing meta-analysis methods such as comparative case studies, forest plots, and scatterplots to ascertain patterns in how a demonstration feature is associated with observed changes in demonstration-related outcomes, for the entire population of demonstration beneficiaries and for subgroups of interest with known health equity gaps.
- Sharing results of the meta-analyses through a series of brief reports and a final report.

SECTION 1.
INTRODUCTION

1.1 Overview of Meta-Analysis Support Contract Goals

The Centers for Medicare & Medicaid Services (CMS) has five objectives for the meta-analyses of section 1115 SMI/SED demonstrations:

1. Use and explore available state and federal data, including the Transformed Medicaid Statistical Information System (T-MSIS) and other sources, to study the effectiveness of Medicaid section 1115 serious mental illness/serious emotional disturbance (SMI/SED) demonstrations and compare the effectiveness of the policies and features across states, including effectiveness for subpopulations of interest that might experience gaps in health equity
2. Provide information, including best practices and recommendations for improving demonstration policy and implementation strategies, to inform national policy making and to support scaling up and diffusion of successful demonstration policy experiments
3. Provide materials for and participate in CMS Learning Collaboratives if requested by CMS
4. Cooperate with CMS, its evaluation and other contractors, and other federal agencies to share data and provide information to improve overall understanding of any related studies
5. Inform CMS on the rigor and limitations of state evaluation designs and reports, as well as monitoring protocols and reports, to support further improvements and capacity building in CMS monitoring and evaluation of section 1115 demonstrations.

This evaluation design focuses on section 1115 demonstrations for SMI/SED.

1.2 Approaches to Delivering SMI/SED Services in Medicaid

States may deliver behavioral health services to individuals with SMI/SED through the Medicaid state plan or Medicaid demonstrations. Both mandatory and optional Medicaid benefits are available to states to deliver mental health treatment services, and the Mental Health Parity and Addiction Equity Act requires that coverage of mental health and substance use disorder services for Medicaid beneficiaries can be no more restrictive than the coverage provided for medical/surgical conditions. In addition to mandatory and optional state plan benefits, states have additional avenues to craft benefit packages to meet the needs of individuals with mental illness. For example, states can apply for a Medicaid Health Home State Plan amendment (SPA), which allows states to design and deliver programs and services that coordinate physical, behavioral health (both mental health and substance abuse), and long-term services and supports for high-

need, high-cost Medicaid beneficiaries. States can design these health home services specifically for individuals with SMI (CMS, 2021b).

States can also utilize a section 1915(c) home and community-based services (HCBS) waiver, which permits states to offer intensive community-based services comparable to an institutional level of care to people who require long-term services and supports (MACPAC, n.d.). These services can target individuals with mental illness, intellectual or developmental disabilities, and physical disabilities. While service categories vary by state, states must provide case management; other HCBS waiver services can include caregiver support; community transition; day services; equipment, technology, and modifications; home-based services; non-medical transportation; other mental health and behavioral health services; and supported employment. Another option to support HCBS service delivery is the 1915(i) SPA, which allows states to set the qualifying level for HCBS at either an institutional level of care or lower without obtaining a waiver from CMS. The 1915(i) SPA permits states to design service packages targeted to people with specific needs, including beneficiaries who have developmental disabilities or mental illness. Examples of commonly offered service categories in 1915(i) mental health SPAs include crisis intervention, behavior support, counseling, and psychosocial rehabilitation.

Targeted demonstrations such as Medicaid's Money Follows the Person (MFP) demonstration are also available to states to support delivery of behavioral health care. MFP provides states with enhanced federal funding to support seniors and people with disabilities who are moving from institutions to the community (Musumeci et al., 2019). Over 80 percent of MFP enrollees are people with physical, mental health, or adult-onset cognitive disabilities.

More recent efforts to meet the needs of individuals with SMI include a provision in the American Rescue Plan Act of 2021 that funds planning grants to states to develop community-based mobile crisis intervention services in Medicaid and provides enhanced Medicaid federal funding to support this new service.

Even though states have considerable flexibility in shaping treatment services for individuals with SMI, in general they cannot receive federal Medicaid funds for services provided to a Medicaid beneficiary between the ages of 21 and 64 who is receiving care in an institution for mental disease (IMD). The section 1115 SMI/SED demonstrations directly address this limitation to Medicaid state plan benefits, the details of which are discussed in ***Section 1.3***.

1.3 Overview of Section 1115 SMI/SED Demonstrations

Improving care for Medicaid-enrolled adults with SMI or youth with SED through innovative service delivery is a priority for CMS. Section 12003 of the 21st Century Cures Act mandated CMS to develop section 1115 demonstration projects for this population. Section 1115 SMI/SED demonstrations allow states to receive federal financial participation (or federal Medicaid matching dollars) for care delivered in an IMD as long as the state is taking action to meet five goals and a series of related milestones to ensure quality of care in IMDs, improve access to mental health care for Medicaid enrollees with SMI or SED across the continuum of care, improve transitions between levels of care, and improve health outcomes. States must also commit to maintaining funding levels for outpatient community-based mental health services and monitor and evaluate demonstration performance.

What Is Serious Mental Illness and Serious Emotional Disturbance?

- **Serious Mental Illness (SMI):** A diagnosable mental, behavioral, or emotional disorder of sufficient duration to meet diagnostic criteria, that has resulted in functional impairment which substantially interferes with or limits one or more major life activities like activities of daily living (e.g., eating, bathing, maintaining a household, getting around the community, taking prescribed medication) and limits functioning in social, family, and vocational/educational contexts.
- **Serious Emotional Disturbance (SED):** A diagnosable mental, behavioral, or emotional disorder of sufficient duration to meet diagnostic criteria that resulted in functional impairment which substantially interferes with or limits the child's role or functioning in family, school, or community activities.

Under the section 1115 SMI/SED demonstrations, states have flexibility in how they define SMI and SED. For example, states can adopt the National Committee for Quality Assurance's definition of SMI, or they can use another definition based on diagnosis and procedure codes of their choosing.

Source: As defined by the Substance Abuse and Mental Health Services Administration and noted in the 11/13/2018 State Medicaid Director Letter, https://www.medicaid.gov/federal-policy-guidance/downloads/smd18011.pdf.

Federal financial participation for IMD stays. Medicaid programs have been statutorily prohibited from using federal Medicaid funds to pay for services rendered to adults aged 21 through 64 in IMDs. Known as the IMD exclusion, this policy was put in place to prevent states from shifting costs for psychiatric institutional care that had traditionally been covered by the states to the federal government. Despite this exclusion, some flexibilities have historically been afforded states to receive federal funds for providing care in an IMD. Coverage is often limited in

- The term IMD only has meaning in the Medicaid program; other payers and accrediting bodies do not use this provider type.
- How states define IMDs varies, but **generally IMDs include psychiatric hospitals and residential treatment settings**.

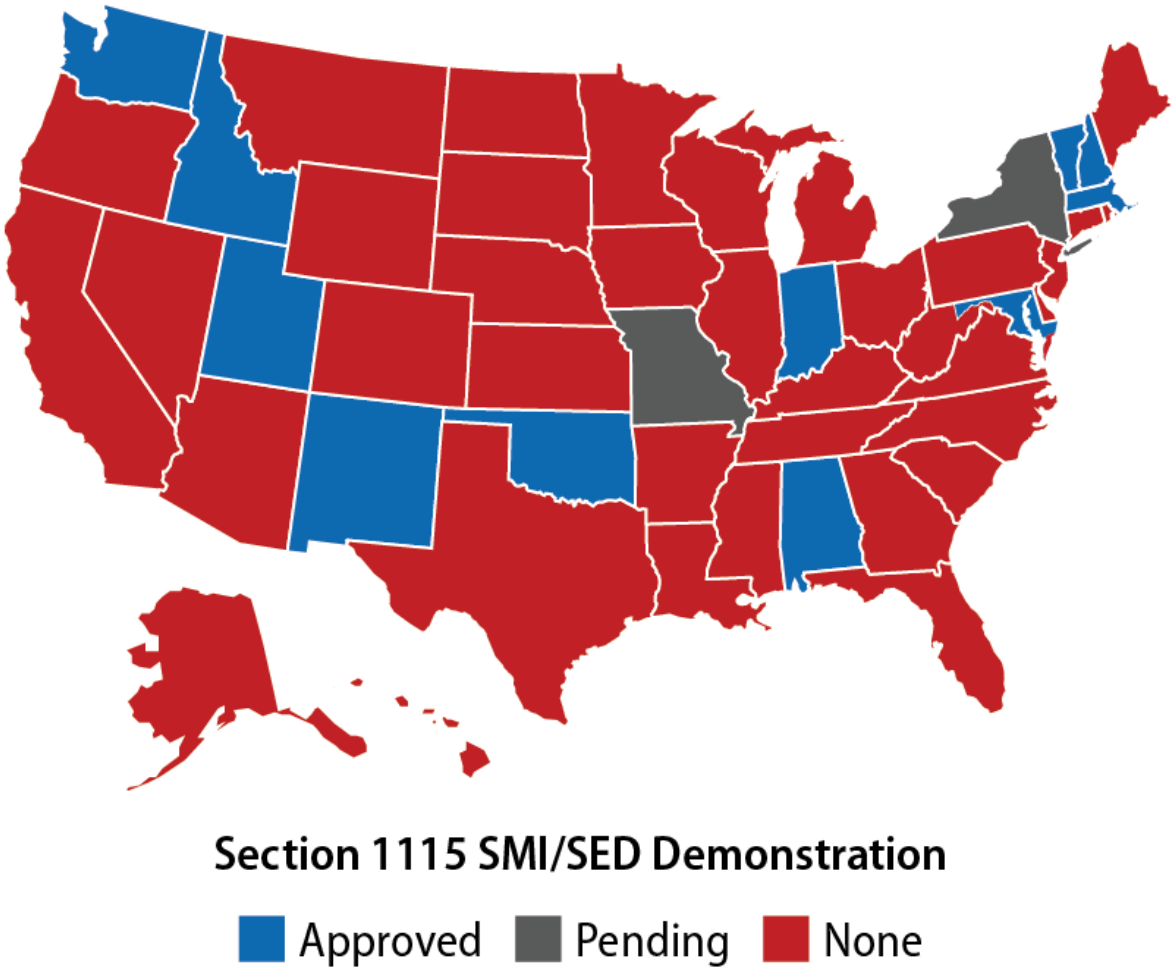
scope and targeted to very select groups. For example, through the section 1115 substance use disorder demonstration, states can provide short-term substance use disorder treatment in IMDs. Through a provision in Medicaid regulations known as "in lieu of" authority, Medicaid managed care organizations can pay for inpatient psychiatric or substance use disorder treatment during short stays at IMDs for nonelderly adults (MACPAC, 2019). The current section 1115 SMI/SED demonstrations allow states to receive federal financial participation for services furnished to Medicaid beneficiaries during ***short-term stays*** for acute care in psychiatric hospitals or residential treatment settings that qualify as a IMD if those states are also taking action through these demonstrations to ensure quality care in IMDs and to improve access to community-based services for beneficiaries with SMI or SED. Participating states are be expected to achieve a statewide average IMD length of stay of 30 days or less. Federal financial participation for long-term IMD stays (longer than 60 days) is not available through this demonstration.

Improving access to services for mental health care. Under demonstration requirements, states are expected to implement activities to improve mental health care and to orient those activities around achieving the following **five goals** (discussed in more detail in ***Section 1.3.1***):

- Reducing utilization and length of stay in emergency departments
- Reducing readmissions to acute care hospitals and residential settings
- Improving availability of crisis stabilization, intensive outpatient, psychiatric hospital, and residential treatment setting services
- Improving access to community-based services and integration of primary and behavioral health care
- Improving care coordination and continuity of care after a hospitalization or residential treatment stay.

As of April 2023, 11 states (Alabama, Idaho, Indiana, Maryland, Massachusetts, New Hampshire, New Mexico, Oklahoma, Utah, Vermont, and Washington) and the District of Columbia had received approval for section 1115 SMI/SED demonstrations; 2 states (Missouri and New York) have a pending application (***Exhibit 1-1***). We will track Medicaid.gov and CMS's 1115 Demonstration Performance Management Database and Analytics System (PMDA) and confer with CMS to stay up-to-date on demonstration approvals.

Exhibit 1-1. Approved and Pending Section 1115 SMI/SED Demonstrations as of April 2023

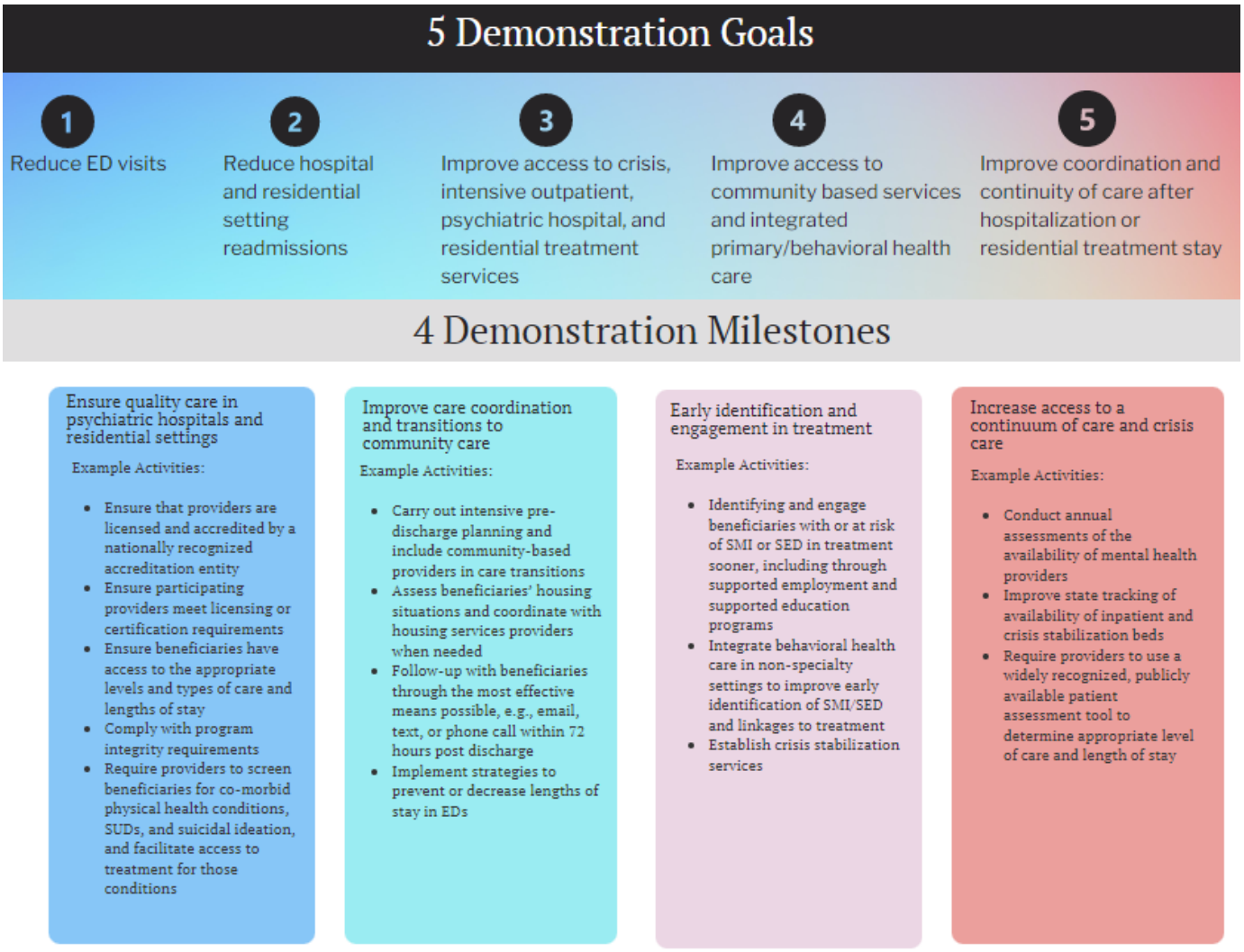


Sources: RTI review of CMS documents, Medicaid.gov, and state documents.

1.3.1 Section 1115 SMI/SED Demonstration Goals and Milestones

To track implementation progress, the section 1115 SMI/SED demonstrations are expected to make progress in meeting the four milestones shown in ***Exhibit 1-2***. As outlined by CMS, achieving these milestones is expected to lead to successful performance on five goals specific to the demonstration. To complete the milestone requirements, states must focus on certain activities, as described in ***Exhibit 1-2***.

Exhibit 1-2. Overview of Section 1115 SMI/SED Demonstration Goals and Milestones



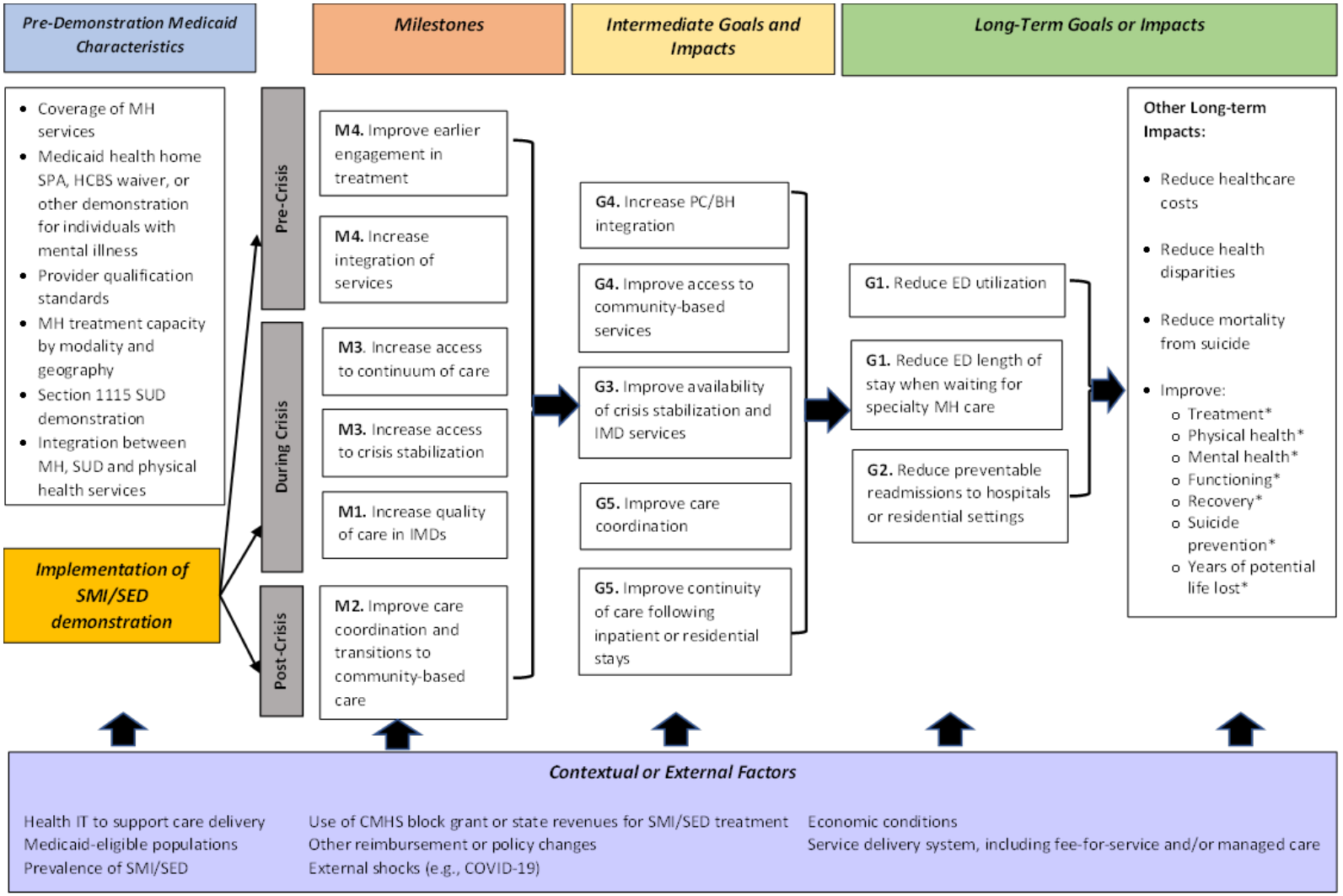
Notes: SUD = substance use disorder; ED = emergency department

1.3.2 Section 1115 SMI/SED Demonstration Conceptual Framework

Exhibit 1-3 provides an overview of the conceptual framework for this meta-analysis of the section 1115 SMI/SED demonstrations.[1] The framework addresses the role of a state's Medicaid benefit package, availability of treatment providers, and approaches to integrating physical and behavioral health care (collectively termed "Pre-Demonstration Medicaid Characteristics") in addition to "Contextual or External Factors" in the design and implementation of a state's demonstration. Through demonstration activities, states are expected to make progress in or meet the four milestones discussed in ***Section 1.3.1***, with the expectation that milestone progress or achievement will lead to demonstration goal achievement, improved health outcomes, reduced health disparities and improved health equity, and reduced cost of care. This framework also situates the demonstration milestones and goals within a continuum of care, from early identification of mental health needs before crisis care is needed through post-crisis care aimed at improving care coordination and care transitions after a mental health-related acute care stay in an IMD or other facility.

[1] The conceptual model presented here for the meta-analysis is not meant to supplant any logic models or driver diagrams created by states and their evaluators for purposes of conducting state-specific monitoring and evaluation of the demonstration.

Exhibit 1-3. Conceptual Model for the Section 1115 SMI/SED Demonstration



Notes: MH = mental health; SUD = substance use disorder; ED = emergency department; IMD = institutions for mental disease; M = SMI demonstration milestone; G = SMI demonstration goal; PC = primary care; BH = behavioral health

*Denotes that the long-term outcome will not be examined in this evaluation due to data availability.

CMS has identified five goals, or desired impacts, for the section 1115 SMI/SED demonstrations. For purposes of this meta-analysis, we will also consider the impact of this demonstration on health care costs for individuals with SMI or SED, thereby adding a sixth goal. Achieving demonstration milestones are expected to impact these goals as follows:

(G1) **Reduced utilization and length of stay in emergency departments among Medicaid beneficiaries with SMI/SED while awaiting mental health treatment in specialized settings**. Increasing the availability of SMI/SED treatment, crisis stabilization services, and integration of behavioral health services with primary care will increase the likelihood that individuals receive appropriate treatment, thereby avoiding inappropriate use of the emergency department.

(G2) **Reduced preventable readmissions to acute care hospitals and residential settings.** Increasing access to appropriate treatment and care coordination services will decrease the rate of relapse after discharge from a treatment episode, which will reduce readmissions to the same or higher level of care.

(G3) **Improved availability of crisis stabilization services, including services made available through call centers and mobile crisis units, intensive outpatient services, as well as services provided during acute short-term stays in residential crisis stabilization programs, psychiatric hospitals, and residential treatment settings throughout the state**. Increasing access to the continuum of care will improve availability of crisis stabilization services as well as more intensive residential or psychiatric hospital services.

(G4) **Improved access to community-based services to address the chronic mental health care needs of beneficiaries with SMI/SED, including through increased integration of primary and behavioral health care.** Increasing access to community-based services will promote early identification and engagement in mental health treatment and better care coordination.

(G5) **Improved care coordination, especially continuity of care in the community following episodes of acute care in hospitals and residential treatment facilities.** Increasing the availability of coordinated SMI/SED treatment and integrated behavioral health and primary care services will improve coordination and continuity of care following episodes of acute care in hospitals or residential facilities.

(G6) **Reduced health care costs for individuals with SMI or SED.** Increasing the availability of coordinated SMI/SED treatment and integrated behavioral health and primary care services may reduce reliance on emergency department, inpatient, or residential care and reduce overall total Medicaid spending.

SECTION 2.
META-EVALUATION ANALYTIC FRAMEWORK, EVALUATION QUESTIONS, EVALUATION OUTCOMES, AND DATA SOURCES

2.1 Section 1115 SMI/SED Demonstration Meta-Analysis Framework

States are required to monitor and evaluate their section 1115 demonstrations per 42 CFR §431.424 and §431.428. Monitoring provides early and ongoing information about demonstration implementation and progress toward goals that CMS and states can use to identify potential problems and to make midcourse adjustments if needed. Evaluations seek to understand demonstration implementation progress and impacts of demonstration activities on health care use, health care costs, beneficiary health, and equity in health and health care. CMS and states can use evaluation findings to make program changes, as needed, to achieve desired outcomes. Evaluations provide evidence to support decisions about whether demonstrations should be extended, whether new demonstrations should be approved, if demonstration requirements should be modified, and, ultimately, whether federal Medicaid policy should be changed.

However, individual, state-specific monitoring reports and evaluations are not well positioned to identify overarching lessons that can be used to shape broader state and federal Medicaid policies to improve care for individuals with SMI/SED. With a meta-analysis, the focus is on identifying patterns in how states implement the demonstration and how successful states are in meeting demonstration milestones and goals. Specific analytic methods are applied to help discern which policies, program components, and contextual factors contribute to meeting demonstration milestones and goals.

Meta-evaluation of cross-state data is:

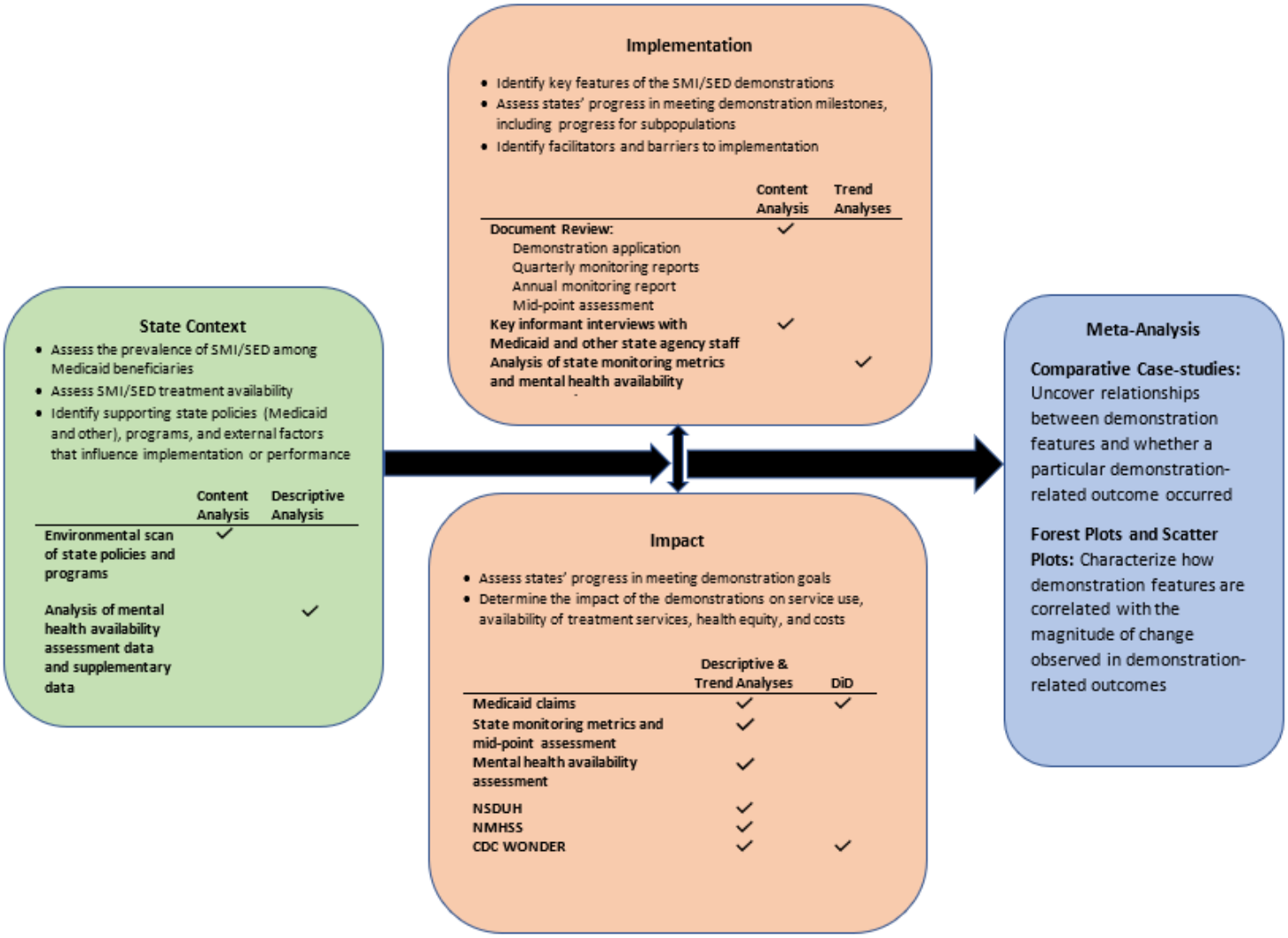
- Concerned with **variations in demonstration design and context** that can impact an individual demonstration's outcomes and explain variation in outcomes observed across demonstrations
- Drawn on **experience across multiple states** implementing differing activities, but with the same policy goal
- Designed to provide CMS and states with a deeper understanding of **what levers affect successful implementation and impacts, and how specific policy initiatives should be replicated** in other states for maximum impact.

Exhibit 2-1 provides an overview of our approach to data collection and analysis to identify patterns in demonstration outcomes and the cross-state factors that may explain those patterns. To characterize the context in which states implemented their demonstrations, we will conduct environmental scans of state policies and programs and analyses of demonstration states' data on the availability of mental health treatment services. To understand exactly what demonstration activities were implemented, on what schedule, if changes were made to planned activities, perspectives on facilitators and barriers to implementation, and if activities led to

changes in demonstration milestones metrics, we will review program documents, conduct key informant interviews, and analyze trends in state monitoring metrics and data on the availability of mental health treatment services. To assess state progress toward meeting demonstration goals, we will conduct quantitative analyses of multiple secondary data sources. Once we can characterize state context, demonstration implementation features, and the impact of individual state demonstrations on milestones and goals, we will apply meta-analysis methodologies to identify factors associated with demonstration-related outcomes across states.

The state-specific implementation evaluation questions are described in ***Section 2.2*** and the cross-state meta-evaluation research questions are described in ***Section 2.3***. A detailed description of the data sources referenced in ***Exhibit 2-1*** can be found in ***Section 2.4***, and a discussion of the analytic methods applied to these data can be found in ***Sections 3.1, 3.2, and 3.3***. Meta-analysis methods are discussed in ***Section 3.3***. All proposed evaluation questions and outcomes are subject to change if CMS identifies emerging priorities or topics for which they would like to have more information or if unanticipated and significant limitations arise in data collection, data quality, or data availability, particularly for state-reported monitoring and outcome data.

Exhibit 2-1. Overview of Approach to Data Collection and Analysis



Notes: SUD = substance use disorder; DiD = Difference-in-differences; NSDUH = National Survey on Drug Use and Health; NMHS = National Mental Health Services Survey; WONDER = Wide-ranging Online Data for Epidemiologic Research.

2.2 Meta-Evaluation Implementation Questions and Outcomes

Exhibit 2-2 summarizes the implementation evaluation questions and the corresponding outcomes. Evaluation questions are organized around the four milestones that states are expected to meet as they roll-out or change demonstration activities. Questions were formulated with the goal of identifying demonstration features and facilitators and barriers to implementation that will support the cross-state meta-analysis evaluation. Evaluation questions and their associated outcomes were chosen with the following considerations:

- There is at least one evaluation question per milestone
- There is at least one evaluation question about the state's health IT plan related to demonstration activities
- There is at least one evaluation question about the state's plans to finance community-based mental health services
- Outcomes can be qualitive or quantitative
- Quantitative outcomes align with section 1115 SMI/SED demonstration monitoring and evaluation metrics (outlined by CMS in monitoring and evaluation guidance) to leverage state-reported data in the meta-evaluation
- Some outcomes reflect states' efforts to drive positive changes for all beneficiaries, thereby improving health equity.

Exhibit 2-2. Implementation Evaluation Questions and Study Outcomes

Milestone 1: **Ensuring Quality of Care in Psychiatric Hospitals and Residential Settings**

Question: What strategies were implemented to ensure quality of care in psychiatric hospital and residential settings?

Outcome:

- State-reported strategies to ensure quality (e.g., utilization review, screen for co-morbid conditions, licensure requirements, program integrity, oversight processes)

Milestone 2: **Improving Care Coordination and Transitioning to Community-Based Care**

Questions: What strategies were implemented to improve care coordination and transitions across the continuum of care and were those strategies effective? What were the facilitators and barriers to implementation of those strategies? Did the strategies impact subpopulations of interest (e.g., children, justice-involved beneficiaries, individuals living in rural versus urban communities) ?

Outcomes:

- Strategies to facilitate pre-discharge planning and post-discharge follow-up
- Strategies to prevent ED stays or reduce ED length of stay
- Strategies to assess housing needs and coordinate with housing services providers
- Strategies to coordinate medical care with behavioral health care

Milestone 3: **Increasing Access to a Continuum of Care, Including Crisis Stabilization Services**

Question: Have states been able to maintain mental health provider availability throughout the demonstration period?

Outcome:

- Counts of mental health providers accepting Medicaid patients, including community mental health centers, authorized prescribers, and practitioners certified to treat mental illness
- Strategies to maintain mental health provider engagement during the COVID-19 pandemic

Questions: What strategies did states adopt to improve tracking of available inpatient and crisis stabilization beds, and what were the facilitators and barriers to implementation of those strategies?

Outcomes:

- Strategies to track psychiatric inpatient and residential setting beds
- Strategies to track crisis stabilization beds

Questions: Did demonstrations implement patient assessment tools, and what were the facilitators and barriers to implementation?

Outcome:

- Use of patient assessment tools

Question: Did the number of beneficiaries with SMI/SED treated in an IMD change after the demonstration began? How did changes in the numbers vary by subpopulations of interest (e.g.,., children, individuals living in rural versus urban communities) ?

Outcome:

- Number of SMI/SED beneficiaries receiving care in an IMD (how the COVID-19 pandemic impacted the number receiving care will also be considered)

Question: Did the number of beneficiaries receiving intensive outpatient and partial hospitalization services, outpatient, telehealth, and crisis stabilization services change after the demonstration began? How did changes in the numbers vary by subpopulations of interest (e.g., children, individuals living in rural versus urban communities)?

Outcome:

- Number of SMI beneficiaries receiving mental health-related intensive outpatient or partial hospitalization services, outpatient services, telehealth, and crisis stabilization services (how the COVID-19 pandemic impacted the number receiving care will also be considered)

Question: Did access to preventive/ambulatory health services for individuals with SMI/SED change after the demonstration began? How did changes in access vary by subpopulations of interest (e.g., children, individuals living in rural versus urban communities)?

Outcome:

- Number of SMI/SED beneficiaries with a preventive/ambulatory health service (how the COVID-19 pandemic impacted the number receiving care will also be considered)

Milestone 4: **Earlier Identification and Engagement in Treatment, Including Through Increased Integration**

Question: How have the demonstrations promoted earlier identification and engagement in treatment, and what were the facilitators and barriers to implementation of those strategies? Did the strategies impact subpopulations (e.g., children, individuals living in rural versus urban communities)?

Outcome:

- Strategies to implement early engagement in treatment

Question: How did demonstration-related activities to promote integration of behavioral health in non-specialty settings improve access to treatment and early identification? Did the strategies impact subpopulations of interest (e.g., children, individuals living in rural versus urban communities)?

Outcome:

- Integration strategies

Question: Did the demonstrations establish more specialized services, including crisis stabilization services for youth and adolescents?

Outcomes:

- Number of crisis call centers, response teams, mobile units, and assessment centers that treat youth
- Strategies to establish crisis services and barriers and facilitators

Other: **Financing and Health IT**

Question: What financing approaches did demonstrations implement to improve availability of community-based mental health care?

Outcome:

- Financing strategies to increase availability of services

Question: Did the demonstrations leverage health IT to improve referrals between providers, electronic care plans, and electronic transitions of care documents, advance care coordination, and alert providers to patients at risk for discontinuing engagement in mental health treatment?

Outcome:

- Health IT strategies newly adopted or extended under the demonstration

2.3 Meta-Evaluation Research Questions and Outcomes

The meta-analysis of state evaluation findings will focus on six specific research questions and ten outcomes. Research questions are organized around the five demonstration goals that states are expected to meet and a sixth goal related to health care costs. The choice of which questions to ask and which outcomes to study were based on the following:

- There is at least one question per demonstration goal
- Outcomes are closely related to the demonstration goal and align with suggested evaluation outcomes detailed in CMS SMI/SED evaluation guidance (e.g., the proposed outcome for the goal of improving care coordination after hospitalization is

a claims-based measure of the percent of mental health admission discharges with a mental health-related follow-up visit within 7 and 30 days)

- Outcomes, when feasible, align with SMI/SED-required monitoring metrics so that state-specific findings are available throughout the demonstration period from quarterly and annual monitoring reports
- Outcomes can be calculated with available Medicaid claims data if state-reported impact results are not available in time to meet the needs of this contract's deliverables
- Outcomes can be examined by subpopulations of interest to better understand disparities in health care and health equity
- Evaluation questions and outcomes are aligned with guidance for the demonstration evaluations states are required to conduct so that impact estimates will be available for meta-analyses from state evaluations.

These evaluation questions and outcomes selected for the meta-analysis are summarized in ***Exhibit 2-3***. For additional information on plans to address the influence of the COVID-19 pandemic on proposed study outcomes, see ***Section 3.5.***

Exhibit 2-3. Meta-Evaluation Research Questions and Study Outcomes

(*Goal 1*) Does the demonstration result in reductions in ED visits among Medicaid beneficiaries with a SMI/SED diagnosis?

- ED visits per 1,000 Medicaid beneficiaries with a SMI/SED diagnosis

(*Goal 2*) Does the demonstration result in reductions in preventable readmissions among Medicaid beneficiaries with a SMI/SED diagnosis?

- 30-day hospital readmissions per 1,000 discharges for Medicaid beneficiaries with a SMI/SED diagnosis

(*Goal 3*) Does the demonstration increase access to and use of intensive mental health services for Medicaid beneficiaries with a SMI/SED diagnosis?

- Number of mental health-related acute hospital admissions and psychiatric hospital admissions per 1,000 Medicaid beneficiaries with a SMI/SED diagnosis
- Number of mental health-related intensive outpatient or partial hospitalization services, IMD stays, and residential setting stays per 1,000 Medicaid beneficiaries with a SMI/SED diagnosis
- Number of crisis stabilization services per 1,000 Medicaid beneficiaries with a SMI/SED diagnosis

(*Goal 4*) Does the demonstration increase access to and use of community-based mental health services for Medicaid beneficiaries with a SMI/SED diagnosis?

- Number of mental health-related outpatient services per 1,000 Medicaid beneficiaries with a SMI/SED diagnosis

(*Goal 5*) Does the demonstration result in improved continuity of care in the community for Medicaid beneficiaries following stays in hospitals and residential treatment facilities?

- Percent of mental health-related admission discharges with a mental health-related follow-up visit within 7 and 30 days
- Percent of mental health-related ED visits with a mental health-related follow-up visit within 7 and 30 days

(*Goal 6*) Does the demonstration change total and mental health-related Medicaid expenditures among Medicaid beneficiaries with a SMI/SED diagnosis?

- Total per beneficiary per month Medicaid expenditures
- Total per beneficiary per month mental health-related Medicaid expenditures
- Ratio of mental health-related non-inpatient and non-residential spending to total mental health-related spending

2.4 Data Sources

2.4.1 Data Sources for State Context, Demonstration Features, and Implementation and Meta-Evaluation Outcomes

2.4.1.1 State Documents

Implementation outcomes and SMI/SED demonstration impacts are hypothesized to be affected by the state's pre-demonstration SMI/SED systems, including their existing approaches to delivering SMI/SED treatment services, the gaps in their benefit packages, availability of qualified providers to render services, and adoption of other standards such as requirements for integrating physical and behavioral health care. These pre-demonstration factors influence states' strategic decisions about which types of programs, policies, and benefits to put in place to meet demonstration milestones and goals. The pre-demonstration context along with the demonstration activities states pursue are collectively referred to as "demonstration features" throughout the remainder of this document.

State documents and interviews are used to identify demonstration features, which include factors that characterize the context in which states are operating their demonstrations and **state actions** to implement programs, policies, and benefits to meet demonstration milestones and goals.

Primary data sources to identify demonstration features include:

- State applications, implementation plans, and evaluation designs
- State quarterly and annual monitoring reports and midpoint assessments of performance
- Interim and summative state evaluation reports
- Stakeholder interview transcripts
- State documents and websites detailing mental health benefits and programs.

Information on demonstration features and how they change over time will be obtained mainly from secondary data sources, including demonstration documents and other state documents. Some examples of these documents include:

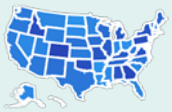
- State applications for the section 1115 SMI/SED demonstration
- Draft and approved state implementation plans
- Draft and approved state demonstration evaluation designs
- State quarterly and annual monitoring reports
- State-reported midpoint assessments
- Interim and summative state evaluation reports
- CMS demonstration approval letters and special terms and conditions (STCs)
- State documents that describe mental health-related benefits and programs for Medicaid beneficiaries (e.g., Medicaid state plan documents and information on state Medicaid and behavioral health agency websites).

2.4.1.2 Stakeholder Interviews

Secondary data will be supplemented with primary data collected through interviews with key stakeholders in demonstration states. Virtual (by phone or video) interviews with stakeholders will be used to confirm information on Medicaid policy changes and other SMI/SED policy changes occurring as a part of the demonstration and to provide a contextual narrative on the impact of the demonstrations. After abstracting demonstration features from the state waiver applications, STCs, implementation plans, and monitoring reports, we will use stakeholder discussions to update and clarify information not covered in the regular reporting or not consistently reported across states. For example, perspectives on the effectiveness of strategies discussed in the implementation plan provide critical context in interpreting states' progress in meeting demonstration milestones and goals. Moreover, interviewees will be able to discuss how demonstration activities impact quality of care for subpopulations of interest, such as youth and adolescents with SMI/SED and individuals residing in urban and rural areas.

Stakeholder interviews are also necessary to accurately document variation among states in demonstration design elements. Information from interviews will be used as inputs for defining the demonstration features used in the meta-analysis and will facilitate a rich interpretation of results from quantitative models. We will also use the interviews to identify barriers, facilitators, and best practices in implementation as state demonstrations evolve.

Virtual Stakeholder Interviews



10 states and the District of Columbia

Interviews conducted at two points: (1) near the midpoint of the demonstration and (2) near the end of the demonstration period

Interviews with Medicaid agency staff, behavioral health agency staff, mental health providers, and social service organization staff partnering with the Medicaid agency

- More than nine individuals will be interviewed across the section 1115 SMI/SED demonstration states, so Office of Management and Budget (OMB) approval is necessary under the Paperwork Reduction Act.

We will conduct interviews with officials from the state Medicaid agency and/or the single state agency for behavioral health during Option Years 3 and 4 of the meta-analysis contract. These two types of stakeholders are in the best position to describe a state's demonstration planning process; demonstration features; the facilitators, barriers, and changes made in implementation; and potential impacts of other initiatives in the state on the demonstration. We may consider expanding the types of interviewees within a state if it makes sense. For example, if a state works closely with organizations providing supportive housing and employment, we may want to interview representatives from one or more of these organizations to gain additional perspectives on how the demonstration is supporting Medicaid members with SMI living in the community. In Option Year 4, we also propose to interview mental health providers, such as large community mental health centers and IMDs that treat Medicaid beneficiaries with SMI. These providers can speak to effective strategies to engage individuals with SMI in care; coordinate care across physical health, mental health, and social service providers; and deliver crisis care.

Preliminary lists of potential interviewees will be shared with CMS, and CMS and RTI will decide which individuals to interview. While we plan to conduct two rounds of interviews, one at a midpoint in each state's demonstration period and one toward the end, we anticipate needing to be flexible in timing to accommodate any delays in program start-up that demonstration states experience and approval of interview protocols in accordance with the Paperwork Reduction Act. If necessary, only one round of interviews could occur in a state.

Exhibit 2-4 describes potential interview topics. Topics of inquiry are likely to shift over the course of the meta-analysis as demonstrations proceed and will be decided in collaboration with CMS. Interview protocols will be tailored to each state and to each stakeholder group within

the state, as well as topics of interest for rapid cycle reporting. Some interview questions will also ask interviewees to expand on discussion of services rendered to subpopulations at risk for disparities in health care and health and strategies for addressing their needs and improving health equity.

Exhibit 2-4. Section 1115 SMI/SED Demonstration Stakeholders Interview Topics

Stakeholder	Potential Interview Topic
State Medicaid agency and/or state behavioral health agency staff	▪ Clarification of information in the implementation plans on major activities and initiatives such as care coordination, care transitions, integration of physical and behavioral health care, and crisis services ▪ Rationale for pursing a section 1115 SMI/SED demonstration ▪ Facilitators and barriers to demonstration activity roll-out ▪ Perceptions of major changes resulting from the demonstration and effects on Medicaid enrollees ▪ Potential impact of other initiatives in the state on the demonstration, e.g., impact of a section 1115 SUD demonstration if the state has one or managed care coverage of IMD services under the "in lieu of" provision in Medicaid managed care regulations
Providers: -Mental health (e.g., large community mental health centers, IMDs) -Social service organizations (e.g., supportive housing and employment providers)	▪ Managing referrals for services for individuals ▪ Strategies for coordinating mental health care, substance use services, and social services across providers ▪ Challenges providing services to individuals with SMI and strategies to overcome those challenges

2.4.1.3 State-Reported Monitoring Metrics

As described in ***Exhibit 2-1***, some implementation outcomes will be assessed using states' quarterly and annual monitoring metrics.[2] With input from subject matter experts and an advisory group, CMS selected 39 metrics to monitor the section 1115 SMI/SED demonstration, of which 31 are required and 8 are recommended. Through monitoring efforts, states and CMS can track demonstration performance and identify areas to adjust to improve performance. We will analyze trends in monitoring metrics as described in ***Section 3.2***.

[2] Monitoring metrics for section 1115 SMI/SED demonstrations are available at https://www.medicaid.gov/medicaid/section-1115-demo/downloads/evaluation-reports/smi-monitoring-metrics.pdf

2.4.1.4 Secondary Data Sources

Seventeen datasets were reviewed, and five national and state-level data sources were selected to help characterize the context in which states implemented their SMI/SED demonstrations and to supplement impact estimates from state evaluations when needed. These data will be used to:

- Determine the prevalence of mental health conditions over time and across states
- Assess SMI/SED treatment availability over time and across states
- Calculate monitoring metrics and impact outcomes that are consistent across states, when needed.

The following datasets will be used to assess mental treatment availability, prevalence of SMI/SED, service use, and health outcomes:

- **Transformed Medicaid Statistical Information System Analytic Files**
  - Service use, mental health treatment availability, and prevalence of SMI/SED in the Medicaid population
- **Centers for Disease Control and Prevention Wide-ranging Online Data for Epidemiologic Research (CDC WONDER)**
  - Health outcomes (i.e., suicide deaths)
- **National Mental Health Services Survey**
  - Mental health treatment availability
- **National Survey on Drug Use and Health**
  - Prevalence of SMI/SED in the Medicaid population and mental health treatment use

Most of the selected datasets contain information on Medicaid enrollees or Medicaid providers, have state identifiers, and have data for all SMI demonstration states approved as of the drafting of this Evaluation Design Report. If information needs shift, additional datasets may be explored or initially selected datasets may not be analyzed. ***Exhibit 2-5*** provides an overview of the datasets, including the years of data expected to be available going back to 2016, the earliest year necessary for analyses.[3] Each data source is discussed in greater detail below.

[3] The first SMI demonstrations were approved in 2019. We propose analyzing data up to three years before a state's demonstration to identify pre-demonstration trends in metrics.

Exhibit 2-5. National Datasets to Be Used in Impact Analyses

Dataset Name	Dataset Description	Years Available for Summative Evaluation Report
Transformed Medicaid Statistical Information System Analytic Files (TAF): Research Identifiable Files (RIFs)	Standardized Medicaid beneficiary enrollment and claims data reported by state Medicaid agencies to CMS and available in the Chronic Conditions Warehouse (CCW). Also include Medicaid provider files. RIFs are updated annually.	Years expected for the summative report: 2016-2021 Currently available: 2016-2021
Centers for Disease Control and Prevention Wide-ranging Online Data for Epidemiologic Research (CDC WONDER)	Contains two mortality databases that capture information on suicide deaths: the Underlying Cause of Death database and the Multiple Cause of Death database. The Underlying Cause of Death database provides mortality and population counts for all U.S. states and counties. Maintained by the CDC.	Years expected for the summative report: 2016-2021 Currently available: 2016-2021
National Mental Health Services Survey	Database of public and private mental health treatment facilities across the U.S., along with information on services offered, facility type, and geographic location. Updated annually. Maintained by SAMSHA.	Years expected for the summative report: 2016-2021 Currently available: 2016-2021
National Survey on Drug Use and Health (NSDUH)	Annual survey that captures prevalence rates of substance use and mental health-related issues and treatment use. Maintained by the Substance Abuse and Mental Health Services Administration (SAMHSA).	Years expected for the summative report: 2016-2021 Currently available: 2016-2021

Centers for Disease Control and Prevention Wide-ranging Online Data for Epidemiologic Research (CDC WONDER)

We will use CDC WONDER **to assess a proposed outcome, death by suicide**—a long-term outcome that may be impacted by SMI/SED demonstration activities.

CDC WONDER contains two mortality databases that capture information on suicide deaths at the state and county level based on death certificates for U.S. residents: the Underlying Cause of Death database and the Multiple Cause of Death database. Information on suicide deaths is not available from other sources, but WONDER has important limitations. First, deaths cannot be limited to Medicaid beneficiaries as source of insurance coverage is not recorded on death data. However, an analysis of county or statewide suicide deaths over time can provide additional understanding of the context in which the demonstrations are occurring, for example, rising suicide rates may be

motivation for a state to focus demonstration activities on expanding crisis services. Second, CDC WONDER suppresses data if there are fewer than 10 mortalities in a state or county and marks rates as "unreliable" when the death count is less than 20; given this, it will be important to assess data availability, particularly at the county level, before making a decision about proceeding with these analyses. Third, the etiology of suicide is multifaceted and complex, and we will not be able to infer correlation between demonstration activities and changes in suicide rates over time.

National Mental Health Services Survey (N-MHSS)

N-MHSS is a publicly available, annual, voluntary survey that collects information on the characteristics of and services offered at public and private mental health treatment facilities across the United States. Detailed information is provided on accepted payers (including Medicaid) and services offered, such as crisis intervention services or other targeted programs for individuals with SMI. Results are not adjusted for facility non-response. Geographic locations of all treatment facilities are provided, allowing for spatial analysis to detect differences in treatment access at the sub-state level.

We will use N-MHSS to assess **mental health treatment availability**, including crisis service and community-based mental health services. Location of facilities offering specialized services can be stratified by urban and rural areas as well as areas of higher and lower social need to assess geographic disparities in treatment availability.

National Survey on Drug Use and Health (NSDUH)

NSDUH is a publicly available, annual survey that collects detailed information on substance use and mental health-related issues among a nationally representative sample of non-institutionalized individuals ages 12 and older in the United States. Survey results can be used to track changes in the prevalence of mental health experiences and select mental health service utilization among individuals who report having Medicaid insurance coverage. Estimates are generated for each state, though the sample size for individual states are small and state-level estimates are pooled across at least 2 years. Thus, it may be difficult to detect change in smaller states.

We will use NSDUH to ascertain **prevalence of SMI** for adults and **major depressive episode** for youth (as a proxy for SED) and to assess **mental health treatment use**. Estimates will be stratified by select groups of interest to assess disparities in prevalence and identify gaps where improvements in health equity may be feasible.

Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files (TAF)

TAF data contain enrollment and claims and encounter information for Medicaid beneficiaries across all states. TAF data include (1) a base summary file that contains beneficiary

enrollment and demographic information, (2) an inpatient file that includes all inpatient facility claims and encounters, (3) an "other therapy" file that contains all professional claims and encounters, (4) a long-term care file that contains all claims and encounters for long-term care, (5) a managed care file that contains information on the managed care organizations, and (6) a provider file that contains provider information. TAF data expand the data that were previously available for Medicaid and include additional data elements, such as expanded beneficiary demographics and managed care plan information.

We will use Transformed Medicaid Statistical Information System (T-MSIS) data to:

- **Identify beneficiaries with SMI/SED** diagnoses using diagnosis and procedure information on claims
- **Identify behavioral health providers** using the provider files
- **Calculate select monitoring metrics and outcomes among Medicaid beneficiaries** with SMI/SED in demonstration states and comparison states over time (see ***Section 3.4.2*** for details about comparison groups)
- **Further understanding of health equity by identifying potential disparities in receipt of health care** by analyzing results by groups of interest.

TAF files are relatively new and are continually updated and improved upon by states and CMS. Therefore, RTI conducted analyses to determine whether the data quality and completeness are sufficient to support analysis of SUD demonstration outcomes. Results from those analyses uncovered several areas of concern for section 1115 SMI/SED demonstration states, including:

- Poor data quality for Utah.
- Inability to identify IMDs because IMD is not a provider type recorded on claims; to identify IMDs, states would need to provide a numerical provider identifier like a national provider identifier or a tax identification number.
- Race data are missing at relatively high rates. However, over the course of this evaluation contract, CMS does plan to introduce the Bayesian Improved Surname Geocoding (BISG) into TAF files. The BISG uses indirect estimation methods to produce probabilistic estimate of race and ethnicity.[4] The BISG will significantly improve our ability to examine select monitoring metrics and outcomes by race and ethnicity.

[4] For more information on BISG, see https://www.rand.org/pubs/periodicals/health-quarterly/issues/v6/n1/16.html.

SECTION 3.
META-ANALYSIS METHODOLOGY

The following sections outline the proposed qualitative, quantitative, and meta-analysis methods to assess individual state progress in meeting section 1115 SMI/SED demonstration goals and to explain cross-state variation in meeting demonstration goals. The proposed methods will also be used to identify existing disparities in access to and quality of care and health outcomes and how the demonstrations' various strategies have supported reducing gaps in health equity.

3.1 Qualitative Analysis of State Documents and Stakeholder Interviews

The primary objective of conducting an environmental scan of state context and analyzing state documents and interview transcripts is to identify a set of demonstration features that will be used in the meta-analysis. Features will be identified and refined as more information on demonstrations becomes available; we will also focus on operationalizing demonstration features that are of highest interest to CMS as they consider policy and program options to improve mental health care for Medicaid beneficiaries.

To identify salient features, we will start with implementation plans and information gleaned from environmental scans. We understand that states may change their demonstration activities, develop new activities not originally detailed in implementation plans, or abandon activities. For this reason, we will rely on quarterly monitoring reports, midpoint assessments, and stakeholder interviews to identify features that are in place at least one year after the demonstration was approved. Examples of possible demonstration features, identified from implementation plans received to date, are described in ***Exhibit 3-1***.

To manage and analyze secondary qualitative data, we will use NVivo 12 software. NVivo is designed for qualitative and mixed methods research and allows integration of other data sources and comparisons within and across demonstrations over time (Bazeley & Richards, 2000; Richards, 2009; Sorensen, 2008).

NVivo facilitates analysis by allowing us to compare and contrast information by research question and by data source or respondent type. We have developed a NVivo 12 database for coding approved SMI/SED implementation plans; coding is organized around current and future demonstration milestones activities in addition to a priori themes that we know will be of interest based upon proposed topics for demonstration implementation case studies (discussed in detail in ***Section 4.1.2***). Example themes include subpopulations like youth, individuals with first-time psychosis, or pregnant women; health equity; responses to the

COVID-19 pandemic; tracking psychiatric inpatient bed availability; and coordination with social services. The database also incorporates state characteristics, for example, if the state has expanded Medicaid, if they have a section 1115 SUD demonstration, and the section 1115 SMI/SED demonstration implementation date. NVivo output will be produced across states and across state characteristics to help us identify emerging themes in the data. The number of state characteristics may increase as a more state demonstrations are approved and we identify additional meaningful characteristics through which to explore themes. The database will also be updated to accommodate coding of stakeholder interview transcripts. We will code implementations plans and other state documents as they become available.

Exhibit 3-1. Examples of Section 1115 SMI/SED Demonstration Features

Demonstration Feature Category	Example Demonstration Feature (Operationalizing the Feature)	
State Context	▪ Expanded Medicaid to individuals up to 138% of the Federal Poverty Level (FPL) (Yes/No) ▪ Medicaid health home SPA for individuals with behavioral health needs (Yes/No)	▪ Prevalence of SMI and SED among the Medicaid population (categorical variable) ▪ Number of treatment providers for SMI/SED prior to the demonstration (categorial variable)
Demonstration Design	▪ Number of years the state implemented demonstration activities (categorical variable) ▪ Coordinated activities with a section 1115 SUD demonstration (Yes/No)	▪ Target population (all individuals with SMI or only those who used an IMD) (categorical variable)
Demonstration Activities	▪ Leveraged managed care to implement activities (Yes/No) ▪ Developed and deployed patient assessment tools (e.g., screeners to assess housing and other social service needs or other tools) (Yes/No) ▪ Changed requirements for follow-up after discharge from an IMD or psychiatric hospital (Yes/No)	▪ Implemented (or expanded) behavioral health care services (Yes/No) ▪ Implemented (or expanded) a tracking system to identify available beds for psychiatric inpatient care or crisis stabilization (Yes/No)

3.2 Analysis of Quarterly and Annual Monitoring Metrics

The primary objective in analyzing state-reported monitoring metrics is to assess how states are doing in implementing the demonstration and meeting demonstration milestones and goals. Monitoring metrics offer an earlier look at milestone progression and demonstration impacts than state evaluation reports, which will not be available for several years. In addition, monitoring metrics assess performance on a broader set of common outcomes than will be possible with the state evaluations.

Exhibits 2-2 and ***2-3*** in ***Sections 2.2*** and ***2.3,*** respectively, summarize the specific monitoring metrics that will be examined in closer detail for this evaluation. Our initial analysis of state monitoring metrics will begin with assessing data quality and completeness among all section 1115 SMI/SED demonstration states. If metrics are not reported consistently across states, TAF data will be used to calculate the metrics that can be derived from claims data. We will then examine (1) trends in metric values over time, (2) trends among subpopulations of interest to assess disparities in health care, and (3) associations between states' performance on monitoring metrics and state demonstration features. In addition to descriptive analyses comparing metrics across groups of states with different demonstration features, the association between state performance on monitoring metrics and demonstration features can be analyzed using comparative case study methods (described in ***Section 3.3.1***).

Monitoring metrics provide an early look at state demonstration performance over time. The following metrics will be assessed with **descriptive trend analyses** and **comparative case study** methods:

- Number of SMI/SED beneficiaries treated in an IMD
- Number of SMI/SED beneficiaries receiving mental health-related intensive outpatient or partial hospitalization services, outpatient services, or telehealth services
- Number of SMI/SED beneficiaries with a preventive/ambulatory health service
- Mental health-related ED visits
- Mental health acute hospital admissions and psychiatric hospital admissions
- Mental health-related intensive outpatient or partial hospitalization services, IMD stays, and residential setting stays
- Crisis stabilization services
- Length of IMD stay
- Medicaid expenditures
- Follow-up after a mental health admission or ED visit
- Medication continuation following inpatient psychiatric discharge

T-MSIS data can be used to supplement and validate state-reported metrics for meta-evaluation.

Analyses of subpopulations of interest. CMS provided guidance on specific populations for which reporting monitoring metrics could be calculated (e.g., age, dually eligible for Medicare

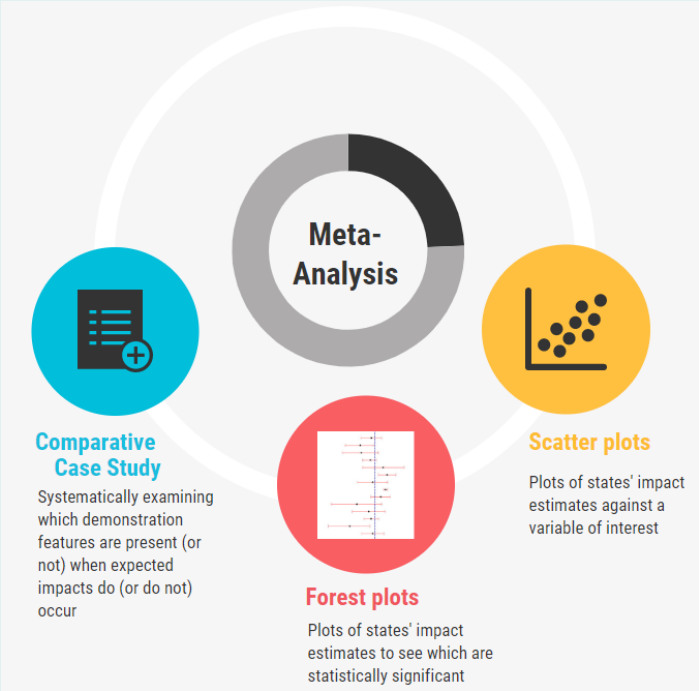
and Medicaid, enrolled in Medicaid due to disability, co-occurring SUD and SMI diagnosis).[5] We will analyze monitoring metrics for subpopulations if our assessment of data quality and completeness shows states report subgroup results consistently and reliably. We will also consult with CMS on priority subpopulations. In the event states do not report metrics by groups of greatest interest to CMS, our team will calculate metrics using TAF data, as described above, so that we can provide results by groups of interest.

Limitations. Analyses of monitoring metrics have some limitations to note. Metrics lack comparison group data because they are intended to be used to understand within-state progress over time. Further, inconsistencies across states in how metrics are calculated may also exist; despite CMS's technical guidance to states, states may deviate from monitoring metric specifications because of data availability or other limitations. States may also vary on how they define the populations for which some monitoring metrics are reported, such as the overall SMI/SED population within the state versus just those who have an IMD stay.

Reporting. Findings from this analysis will be included in the Rapid Cycle Report: Interim Performance Report discussed in ***Section 4.1.1*** and the Summative Evaluation Report.

3.3 Meta-Analysis of State Evaluation Findings

The meta-analysis of state evaluation findings will focus on the six specific research questions and ten outcomes described in ***Section 2.3***. To answer these evaluation questions, we will undertake meta-analyses that will include comparative case studies, forest plots, and scatterplots.



Forest plots and scatter plots and **qualitative comparative case studies** will be used to compare state performance and elucidate how the presence or absence of a feature is associated with a change in evaluation outcomes.

[5] Monitoring metrics for section 1115 SMI/SED demonstrations are available at https://www.medicaid.gov/medicaid/section-1115-demo/downloads/evaluation-reports/smi-monitoring-metrics.pdf

Per demonstration special terms and conditions, each state demonstration must conduct statistical analyses of demonstration effects to be reported in midpoint assessments and final evaluation reports. Midpoint assessment guidance recommends states calculate changes in select monitoring metrics from a baseline period to the midpoint of the demonstration period. The final evaluation guidance recommends that states use a regression-adjusted difference-in-differences model to estimate demonstration effects. The resulting effects are adjusted demonstration versus non-demonstration state values.

These state-specific effects from the midpoint assessments and evaluations are envisioned as the inputs for the meta-analysis of demonstration impacts. Results from the evaluation will be given higher priority as they are expected to reflect more robust analytic methods that support inference about the impact of the demonstration on beneficiaries. However, we may find that midpoint assessment results could be used in meta-analyses; decisions about using available midpoint assessment data will be made in partnership with CMS. Also, we are prepared to generate the demonstration-specific impacts for meta-analyses ourselves if state evaluation results are not available in time, outcomes are not measured comparably across state evaluations, or they do not use rigorous evaluation designs (for example, the evaluation lacks a comparison group). There is a high likelihood that we will need to generate our own impact estimates. Final summative state evaluations for the first three section 1115 SMI/SED demonstrations approved by CMS (the District of Columbia, Indiana, and Vermont) are not due to CMS until June 2023, which is after the date by which RTI must submit the draft final summative evaluation report to CMS. The remaining section 1115 SMI/SED demonstration states have reports due after June 2023. Even though interim evaluation reports will be due to CMS sooner (one year prior to the expiration of the demonstration approval period), most are not due until 2022 or later; factoring in the time CMS will need to review these interim evaluation reports, we expect approved results will not be available until close to the end of the contract period. Given this timing, we expect we will need to calculate the outcomes for meta-analyses using T-MSIS data. Details of how impact estimates will be generated are discussed in ***Section 3.4***.

Appendix A: Select State Evaluation Plans summarizes several details about evaluation design plans for the District of Columbia, Idaho, Indiana, and Washington.

The states differ in their definitions of the target population and proposed quantitative and qualitative analyses.

As additional demonstrations are approved and implementation plans and evaluation design plans become available, the meta-analysis design may be revised to account for significant changes in the number of demonstration states and evaluation approaches.

3.3.1 Comparative Case Study Analysis

Comparative case study analysis helps to uncover relationships between demonstration features and demonstration outcomes. It is a data analysis methodology that systematically examines similarities and differences across cases on a number of different case study constructs (Goodrick, 2014). For our analyses, state demonstrations will serve as the cases and demonstration features (discussed in ***Section 3.1***) will serve as case study constructs. Demonstration outcomes are characterized in a categorial manner, for example, whether improvements in the outcome were achieved or not. It is important to note that if all states show no change in a particular impact estimate of interest, there is no cross-demonstration variation to explain using meta-analysis methods.

As a first step, we will decide which demonstration features are most relevant for the meta-analysis evaluation question and outcome under study and then further refine the list of potential features to those that vary across the demonstration states; example demonstration features are discussed in ***Exhibit 3-2***. Selected features should have a hypothetical relationship to the outcome, that is, presence or absence of the feature might be expected to impact the outcome positively or negatively. There also needs to be variation across states in the feature; if all states have the feature, we cannot conclude if its presence was correlated with outcomes. For example, if all demonstration states expanded Medicaid to individuals with incomes up to 138 percent of the Federal Poverty Level (FPL), performance on an outcome cannot be explained by expanded Medicaid coverage.

Once the most critical constructs are identified for particular meta-analysis evaluation questions, we will generate a case study comparison grid similar to the example in ***Exhibit 3-2***. Our team will use the comparison grid to systematically identify whether relationships or patterns exist between the presence or absence of specific demonstration features and observed impacts.

Exhibit 3-2. Example of a Comparative Case Study Grid for Section 1115 SMI/SED Demonstrations

State Demonstration Features	Reduction in Mental Health-Related ED Visits				
	Yes			No	
	State A	State B	State C	State D	State E
Medicaid expansion status	●	●	●	○	
Demonstration period longer than 2 years	●	●			
Use managed care contracts to implement demonstration activities	●	●	○	●	○
Changed requirements for follow-up after discharge from an IMD or psychiatric hospital	●	○	●	○	
Developed and deployed a screening tool to assess housing and other social service needs	●	○	●		

● = Demonstration feature is present in this state at least one year after CMS approved the demonstration;
○ = feature is partially present in the state at least one year after CMS approved the demonstration;
empty cell = feature is not present in this state.

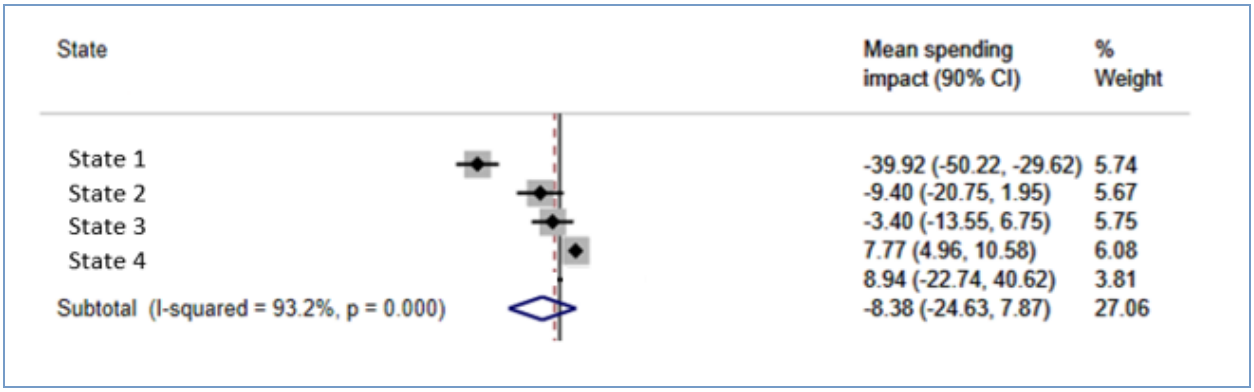
3.3.2 Forest Plots and Scatter Plots of Impact Estimates

Forest plots and scatter plots extend the analysis provided by a comparative case study to incorporate actual values of impact estimates in order to ascertain patterns in how a demonstration feature is associated with the magnitude of change observed in the impact estimate.

Forest Plots. Forest plots display state evaluation results for an outcome by a set of categorical features (i.e., yes, feature is present; no, it is not present). Graphical representation provides a high-level sense of the difference in state results on a particular outcome by a particular feature. To create these forest plots, we gather impact estimates and standardize them so they can be compared. We will then calculate a cross-state estimate of the overall demonstration effect. This mean value is derived by weighting the individual state effect sizes by the precision of each estimate so that more precise estimates from larger programs are given greater weight. The weighted mean will be tested to determine whether it differs significantly from zero. Results are then displayed in a manner similar to ***Exhibit 3-3***. The dot depicts the estimated impact derived from regression-adjusted difference-in-differences models. The horizontal lines show the 90 percent confidence interval (CI) for the estimate. The dotted vertical line demarcates an impact of zero. In this example, dots to left of the vertical line represent favorable changes relative to a comparison group (relative decrease in expenditures); dots to the right indicate unfavorable changes relative to a comparison group (relative increase in expenditures). Results in a plot can be grouped by presence or absence of a feature to facilitate understanding of the relationship

between demonstration features and impact outcomes. Our experience has been that multisite meta-analyses typically yield the general pattern shown in ***Exhibit 3-3*** in which most interventions have insignificant, near-zero impacts with smaller numbers of successful and unsuccessful programs. We expect to see a similar pattern among the section 1115 SMI/SED demonstrations.

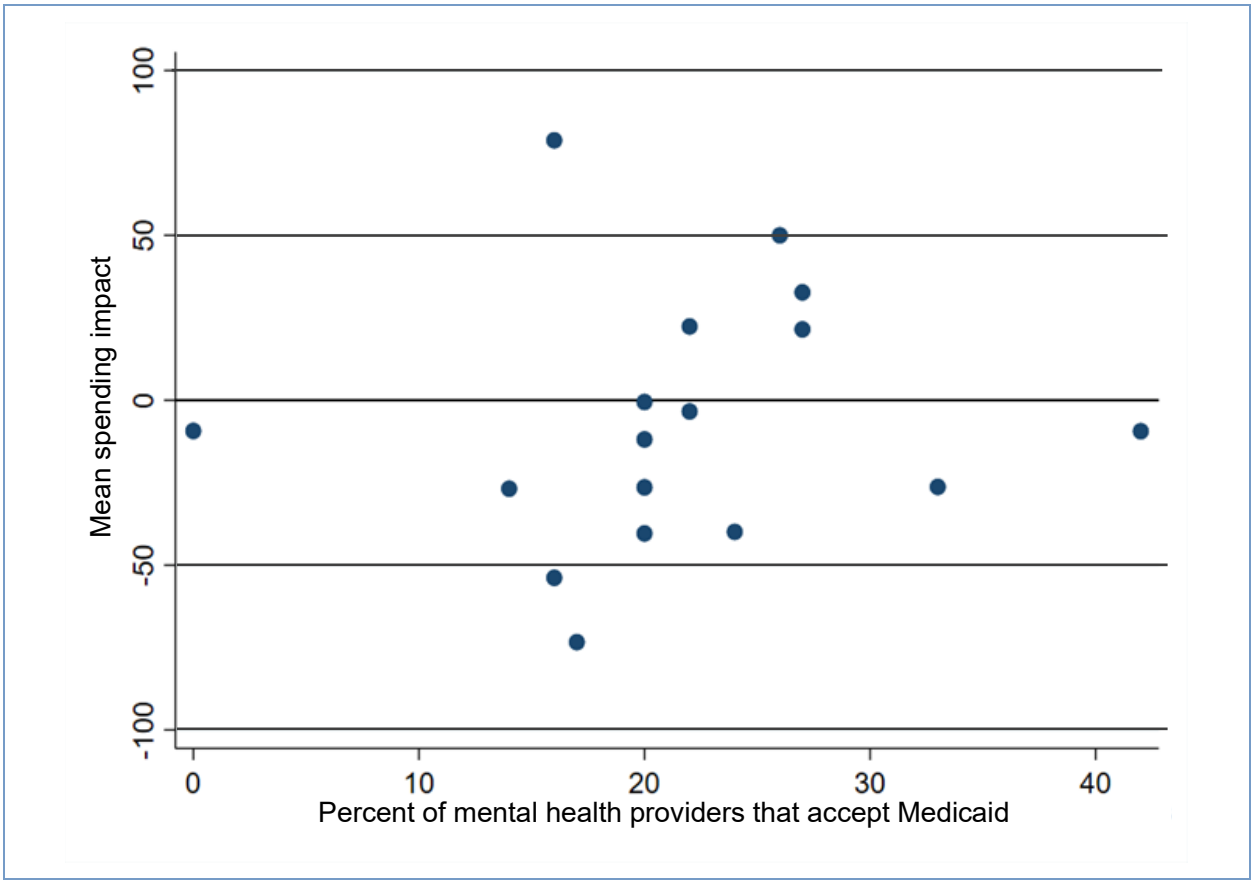
Exhibit 3-3. Example Forest Plot Showing State Demonstration Effect Sizes and Standard Errors for Estimated Medicaid Savings



Scatter plots. With scatter plots, standardized impact estimates are plotted against a variable of interest. The resulting scatter plot is then used to identify a relationship between continuous or categorical demonstration features and a continuous outcome (e.g., number of mental health treatment providers and total Medicaid spending impact estimates). For example, in ***Exhibit 3-4***, impact effect sizes are plotted against a categorical demonstration feature to determine if there is an obvious pattern among states that have that demonstration feature. In this particular plot there is no clear relationship between number of providers and changes in spending.

In addition to showing patterns, the dots in a scatter plot can be used to highlight information that may be useful when interpreting findings. For example, larger dots can be used to signify results that may be considered more reliable because the impact estimates were derived from a larger sample size or more robust evaluation design. Dots can be color coded to show if the impact estimate was statistically significant, or bars can be added to the dots to show confidence intervals around the impact estimates.

Exhibit 3-4. Example Scatter Plot Showing State Demonstration Effect Sizes



Analysis to Examine Disparities in Demonstration Impacts for Subpopulations of Interest. These visual, descriptive analyses can be stratified by subpopulations of interest. For example, results can be stratified by adults with SMI and children with SED or by race, ethnicity, sex, urban/rural area of residence, and Medicaid eligibility category. The nature of demonstration activities, CMS's request for analyses for particular subpopulations, and data availability will guide when and if it makes sense to explore impact results by different groups.

Reporting. We do not expect the demonstration effect results from state's interim summative evaluations to be available until late in the contract period, and if we calculate impact outcomes in lieu of using states' evaluation reports, we will need to allow for sufficient demonstration time before assessing outcome impacts. Therefore, this analysis will be conducted for and presented in the summative evaluation report (see ***Section 4.2*** for more details on the summative evaluation report).

Limitations. Comparative case studies, forest plots, and scatter plots do not quantify how strongly a particular feature is associated with an outcome; that is, there is no numerical measure

of effect. However, not being able to quantify a relationship will not hinder our ability to draw conclusions about how the presence or absence of a feature is associated with a change in study outcomes, and these are the conclusions that will allow CMS to better understand the conditions under which states may be likely to have success meeting demonstration goals.

3.3.3 Differences in the Meta-Analysis Approaches for SMI/SED and SUD Demonstrations

The SUD (substance use disorder) demonstration Evaluation Design Report discussed two additional meta-analysis methods that are not included here, meta-regression and qualitative comparative analysis (QCA). Meta-regression is similar to a standard regression model that aims to quantify the association between variations in demonstration features and the magnitude of an impact estimate. However, to meet statistical assumptions for the validity of the regression model, roughly seven to ten observations are needed for each variable (in this case, each demonstration feature) to be included. With seven SMI demonstration states approved as of June 2021 and an additional five to seven states expected to receive approval in the next 12 months, we do not anticipate having enough states to conduct meta-regression.

QCA is case-oriented approach that examines relationships between explanatory factors (called "conditions" in QCA) and an outcome using set theory, which is a branch of mathematics and symbolic logic that deals with the nature and relations of sets (Schneider & Wagemann, 2012). QCA can be used with small to medium-N studies (i.e., 10 to 100 cases) and is useful for understanding causally complex phenomena. Because ten cases (in this case ten demonstrations) is the bare minimum needed for conducting a robust QCA, we do not anticipate having enough demonstrations to pursue this methodology by the time this contract ends.

3.4 Generating Demonstration Impact Findings in Lieu of Using States' Evaluation Findings

As previously discussed, we expect that state-reported monitoring metrics and state-specific effects from the independent state evaluations will be used to assess state performance over time in meeting milestones and goals and to conduct meta-analyses of cross-state performance. However, we anticipate challenges obtaining and using these data, including:

- Delays in reporting quarterly and annual monitoring metrics
- Inconsistencies across states in how metrics and estimates are calculated (even though CMS has provided technical specifications for monitoring metrics, states may deviate from the specifications when necessary because of data availability or data limitations within their own data systems)
- Timing of state interim and final summative evaluation report submission dates relative to the delivery date for the summative evaluation report from this contract

- Lack of analyses for subpopulations that may be of particular interest to CMS
- Possibility that independent state evaluations may use weaker evaluation designs to produce impact estimates.

We propose difference-in-differences (DiD) analyses to generate impact estimates for each demonstration. The DiD approach is discussed in ***Section 3.4.1***. The amount of time available to calculate impact estimates after the demonstration begins will vary based on when each state began implementing demonstration activities. We anticipate having at least 1-2 years of post-demonstration implementation time to include in DiD models. Because the full impacts of activities may take longer than 1-2 years to materialize, these impact estimates should be interpreted as early impacts.

The unit of analysis for a DiD analysis will be at the beneficiary level because all meta-evaluation impact outcomes can be assessed at the beneficiary level. For example, total costs of care or mental health-related costs of care can be generated at the beneficiary level using Medicaid TAF claims data.

3.4.1 Overview of the DiD Approach

DiD assesses the decline (or growth) in an outcome before and after the demonstration relative to the decline (or growth) in an outcome in a comparison group. ***Equation 1*** shows the basic DiD specification:

$$\text{Outcome}_{i,t,b} = f(\beta_0 + D_i\beta + \Sigma_t \text{Year}_{t,b}\beta_t + \Sigma_k \text{Year}_{k,p}\beta_k + \Sigma_k (D_i * \text{Year}_{k,p})\delta_k + X_{ijt}\theta + \varepsilon_{it}) \quad (1)$$

$Outcome_{i,t,b}$ is a dependent variable indicating the outcome of interest for individual i in a particular time period. D_i (=0,1) is an indicator variable for being in the demonstration group or the comparison group. $Year_{t,b}$ is a series of dummy variables for the baseline measurement years. The subscript t denotes year t out of b baseline years. $Year_{k,p}$ is a series of dummy variables for the demonstration measurement years. The subscript k denotes year k out of p demonstration years. X_{ijt} denotes the set of model covariates; the vector X contains beneficiary- or county-level characteristics, examples of which are described in ***Exhibit 3-5***. ε_{it} is an error term. The coefficient δ_k is the DiD estimate, measuring the impact of being exposed to the demonstration. It represents the average change over time in the outcome for the demonstration group relative to the average change over time in the outcome for the comparison group.

DiD regression models can be used for linear outcomes (e.g., costs), count outcomes (e.g., number of ED visits), and binary outcomes (e.g., had a follow-up visit with a mental health

practitioner after a mental health admission). The DiD model can also accommodate weighting, such that regression-based impact estimates are weighted so that the experience of certain comparison group members are given more consideration in generating the impact estimate than others.

Exhibit 3-5. Potential Beneficiary- or County-level Characteristics to Include in a DiD Regression Model

Beneficiary Level	County Level
Age	Hospital beds per 100,000 population
Sex	Physicians per 100,000 population
Chronic Illness and Disability Payment System (CDPS) risk score	Population age 12-64 covered by Medicaid
Reason for Medicaid enrollment	Metropolitan status
Dually enrolled in Medicaid and Medicare	Population White and Non-Hispanic
	Unemployment rate

DiD assumptions. DiD models assume that the outcomes for the demonstration and comparison groups follow a similar growth trend during the baseline period. Both subjective, visual assessments and statistical tests will be used to check this assumption. Visual inspection of trend graphs of outcome estimates during the baseline period will give a general impression of whether the outcomes were following a similar trajectory during baseline. To complement this visual inspection, we will estimate a regression model with a linear time trend during the baseline period. We test whether this trend differs for the demonstration and comparison group. If there is a difference, we will include a linear time trend variable in the DiD models to correct for divergent baseline trends.

Baseline period. The first several section 1115 SMI/SED demonstrations (Indiana, District of Columbia, and Vermont) were approved in 2019, and their implementation plans were approved concurrently with demonstration approval. We propose analyzing data up to three years before a state's demonstration start to identify pre-demonstration trends in outcomes, and because the first demonstrations were approved in 2019, we will select 2016 as the earliest baseline year. In our previous evaluation work for CMS, we have found that three years of baseline data is sufficient to provide an understanding of baseline trends before a program or policy begins. Moreover, Medicaid claims from TAF data—a key data source for calculating outcomes of interest—are only available back to 2016.

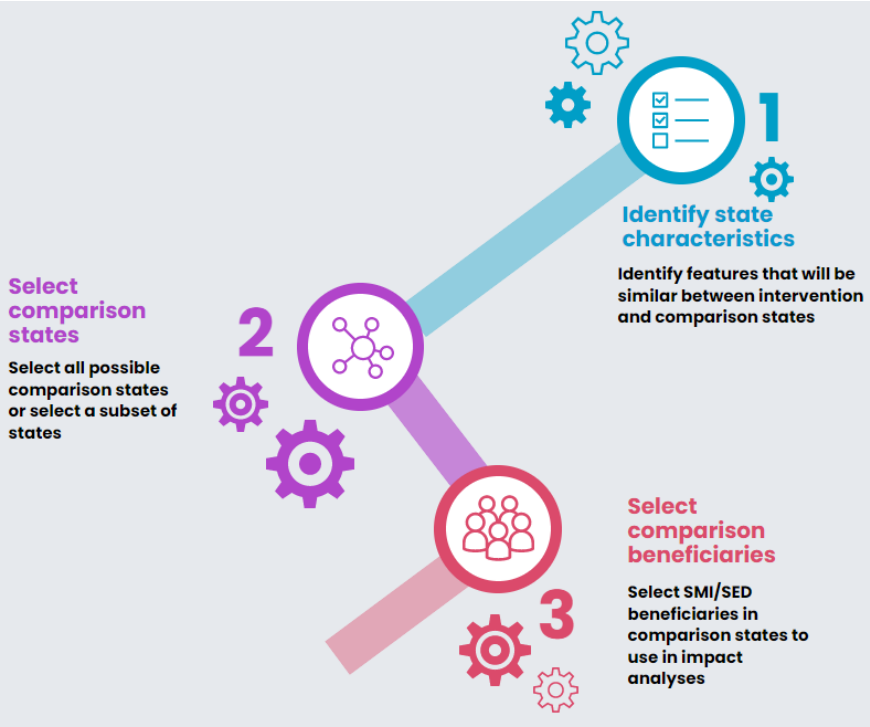
Analyses of subpopulations. The DiD model can also accommodate analyses of subpopulations of interest (e.g., children, individuals with SMI residing in urban vs. rural areas), using stratified analyses (which compare the impact estimates from DiD analyses conducted separately for each group of interest) or triple difference models (which include an interaction between the DiD variable and an indicator for a group of interest). We will consult with CMS on priority populations for which they would like to see impact estimates.

3.4.2 Comparison Group Identification

State evaluations are strongly encouraged to include a comparison group to assess demonstration impacts on outcomes, so for each demonstration state, we will create a comparison group. The comparison group is needed to implement the DiD regressions that generate the impact estimates that will be included in the meta-analyses discussed in ***Section 3***. A within-state comparison group will not be appropriate for most demonstrations because the demonstration aims to address the needs of all Medicaid enrollees with SMI/SED within the state. Therefore, the comparison group will be drawn from other states. If we experience challenges selecting comparison states, we will consider drawing a comparison group from counties in states not participating in the demonstration, after consultation with CMS.

We outline our general approach to selecting a comparison group below. When we are ready to begin these analyses and we know how many and for which states we need to create impact estimates, we will engage CMS on these considerations and prepare a memorandum detailing the proposed approach to selecting a comparison group.

Step 1. Identify characteristics that would be used in selecting comparison states. In a DiD framework, the comparison group should be as similar as possible to the demonstration group. States that look very different from a demonstration state (e.g., sociodemographic makeup of the Medicaid population; comprehensiveness of Medicaid coverage policies) will not make good comparators. Examples of characteristics that will be considered include Medicaid program characteristics (e.g., eligibility levels; use of a Medicaid health home for individuals with mental illness; use of managed care, fee-for-service or both to deliver mental health services), number of Medicaid-enrolled behavioral health providers (e.g., number of IMDs and number of mental health practitioners), prevalence of SMI/SED among the Medicaid-enrolled population, and geographic proximity to the demonstration state. This list will be refined in partnership with CMS when we begin impact analyses.



Step 2. Select comparison states. Depending on the number in the potential pool, all states could be used, or a selection could be made. We expect to have multiple states serving as the comparison group for each section 1115 SMI/SED demonstration state. One common approach to selecting a subset of states is to use a distance score, which is a summary measure of the difference, or "distance," between the characteristics of a demonstration state and a potential comparison state. Distance scores essentially quantify the distance between the demonstration and comparison options along the variables chosen. They make it possible to select comparison states with the nearest match on the characteristics identified in Step 1. There are several unique considerations in this step for the section 1115 SMI/SED demonstration evaluations. First, states are approved over time to implement the demonstration (i.e., rolling-entry of states), and a state that may at first appear to be a good match for an earlier implementing demonstration state may become a demonstration state later. These later entry states will need to be excluded from the comparison group. Second, distance scores may need to be adjusted so that we select enough comparison group states to ensure an adequate sample individuals with SMI/SED. While there is no a priori required number of individuals with SMI/SED for the comparison group, we do want sufficient sample size to support regression-based impact analyses.

Step 3. Select comparison beneficiaries. Once comparison states are selected, beneficiaries with SMI/SED will be selected. Based on the number of individuals with SMI/SED in the comparison states, we may choose to select all individuals for analysis, or we could select a subset. Given that the prevalence of SMI/SED (identified through secondary data like claims data) can be fairly low in some states, we may want to select all individuals available. However, we can give more weight in impact analyses to comparators who look most like individuals in

the demonstration group. One way to do this is to use propensity score models. In a propensity score model, a logistic regression is used to model the probability (or propensity) that an individual is in the demonstration group given a set of characteristics. Then, individuals in the comparison group are assigned a weight in the DiD regression model based on their propensity score and individuals with a higher propensity are given more weight.

Possible covariates in a propensity score model:

- Age
- Sex
- Reason for Medicaid enrollment
- Chronic Illness and Disability Payment System (CDPS) risk score
- Dually enrolled in Medicare and Medicaid.

3.5 Impact of COVID-19 on Meta-Analysis of Implementation and Impact

In January 2020, the federal government declared a public health emergency (PHE) in response to the COVID-19 epidemic, and in March 2020 the national emergency in response to the epidemic was declared. To support Medicaid programs during this time, the federal government increased its share of Medicaid costs; put in place key protections for beneficiaries, such as prohibiting states from cutting Medicaid eligibility while receiving additional funds or for terminating most Medicaid enrollee's Medicaid coverage unless the enrollee specifically requests termination of Medicaid coverage, moves out of state, or dies; and approved numerous section 1135 waivers, section 1115 waivers, and Medicaid and Children's Health Insurance Plan state plan amendments to grant flexibilities to Medicaid providers and state Medicaid agency program operations. For the purposes of this evaluation, we can consider the pandemic period to extend from March 2020—when the national emergency response began—to June 2021, when national mask mandates were lifted. However, a final decision about the length of the pandemic period will be made in partnership with CMS.

To develop a plan to account for this unprecedented and unexpected period into the meta-evaluation of the section 1115 SMI/SED demonstrations, we first propose a set of exploratory qualitative and quantitative analyses to understand provider and beneficiary responses to the PHE (***Exhibit 3-6***). The exploratory analyses are meant to ensure a complete understanding of when demonstration activities began relative to the PHE, how the states had to pivot in program design and roll-out during the PHE, how providers involved in delivering demonstration services had to change service delivery, how prevalent COVID-19 diagnoses were among the target populations, and how beneficiary's use of demonstration-related services changed during the PHE period.

Exhibit 3-6. Stakeholder's Perspective on Provider and Beneficiary Response to the COVID-19 Pandemic

Stakeholder's Perspective on Provider Response to the Pandemic		Stakeholder's Perspective on Beneficiary Response to the Pandemic	
Qualitative Assessment: Implementation Changes	**Quantitative Assessment: Exploratory Analyses**	**Qualitative Assessment: Implementation Changes**	**Quantitative Assessment: Exploratory Analyses**
Stakeholder interviews and information shared in monitoring reports: ▪ Did the state need to make changes in program design or program roll-out because of the pandemic? ▪ How has the pandemic impacted: – The number of providers delivering services? – The availability of demonstration programs or activities (e.g., crisis service pilot programs)?	Use secondary data sources to characterize how the following changed during the pandemic: ▪ The number of providers delivering services. ▪ The availability of demonstration programs or activities (e.g., crisis service pilot programs).	Stakeholder interviews and information shared in monitoring reports to assess how the following changed during the pandemic: ▪ The types and numbers of beneficiaries served through demonstration activities. ▪ The types of services used by beneficiaries (e.g., more telehealth).	Use secondary data sources to characterize the following during the pandemic: ▪ Prevalence of COVID-19 diagnoses in TAF claims data among individuals with SMI or SED. ▪ Trends in utilization and costs before, during, and after the PHE.
Considerations: *When the state was approved and began delivering demonstration services, as defined by approval to receive federal financial participation for short-term IMD stays, relative to the PHE period*: ▪ *Demonstration was approved before the PHE began or just as the PHE began:* District of Columbia, Indiana, and Vermont demonstrations were approved in late 2019 or January 2020, roughly 3 months before the PHE began. Idaho was approved at the start of the PHE in April 2020 ▪ *Demonstration was approved in the middle of the PHE*: Alabama, Maryland, New Hampshire, Oklahoma, Utah, and Washington were approved in late 2020, 2021, or 2022.			

Results of the exploratory analyses will be used to develop adjustments to the analyses for generating state-specific impact estimates to account for COVID-19 impacts.

Implications for state-specific impact estimates. The pandemic will influence state-specific estimates of demonstration impact in several ways. We describe these implications and strategies to adjust for them.

COVID-19 has changed trends in health service utilization. Early evidence examining the impact of COVID-19 on Medicaid enrollees using Medicaid claims data through October 2020 showed notable changes in service use (CMS, 2021a). Specifically, telehealth use increased substantially over the period March through October 2020 relative to the same time a year earlier (March through October 2019), with a peak in telehealth service use through April 2020 and declining rates thereafter. Mental health and substance use services have decreased, 22 percent and 13 percent respectively, for adults 19 to 64 years of age relative to the period March through October 2019, while at the same time some data suggest there has been an increase in the number of adults reporting adverse mental health conditions during the PHE. Our regression models could account for trends like these in several ways, such as excluding particular time periods from a model, adding specific time fixed effects, or defining cohorts of individuals with certain patterns of service use and controlling for those cohorts.

Costs to treat COVID-19 could influence total Medicaid spending. Over the period March through October 2020, an estimated 1.2 million Medicaid and CHIP beneficiaries received treatment for COVID-19 and almost 124,000 were hospitalized (CMS, 2021a). Individuals admitted to the hospital with COVID-19 are expected to have higher costs associated with that episode of care. If the prevalence of the COVID-19–diagnosed population admitted for inpatient care is high within our SMI/SED target population, Medicaid expenditures will be higher than we would typically expect to see. Expenditures may need to be adjusted, for example by removing the COVID-19–related episode costs, to remove the influence that COVID-19–related costs could have on regression-based estimates of impact.

Data on state- and county-level COVID-19 impact and government response will be obtained from:

- The University of Oxford's COVID-19 Government Response Tracker
- Johns Hopkins University Coronavirus Resource Center
- The Centers for Disease Control and Prevention COVID Data Tracker

Geographic variation in COVID-19 response influences selection of a comparison group. Geographic differences in the impacts of the PHE and responses to the PHE will impact how we select a comparison group of states/counties/beneficiaries for a demonstration state. Ideally, comparison states should look similar to a demonstration state in terms of prevalence of COVID-19 diagnoses, COVID-19 death rates, community risk for COVID-19 transmission, hospital resource use, and stringency of state or county government response (i.e. openings and closings of businesses and mask mandates), so these state-/county-level factors will need to be considered when selecting comparators. Similarly, for analyses at the beneficiary level (e.g., calculating service use and expenditure metrics using TAF data), it will be important to create beneficiary-

level indicators of having had a COVID-19 diagnosis so that we can control for or match beneficiaries on having had COVID-19.

As we move forward with exploring the impacts of the PHE in our proposed data sources, we will share results of analyses with CMS and propose plans for addressing what we see in the data. For example, if we see a high proportion of SMI/SED beneficiaries with COVID-19 diagnoses, we may suggest stratified analyses to determine impact estimates for key outcomes like total costs of care and ED visits among those with and without COVID-19. If we see big shifts in service utilization, we may suggest statistical controls in regression models to control for the shifts. If we see a trend toward more-than-expected outlier costs of care due to COVID-19 episodes of care for beneficiaries with SMI/SED, we may need to adjust expenditure estimates to remove COVID-19 related expenditures.

Cross-state meta-analyses. As shown in ***Exhibit 3-6***, states can be grouped by demonstration approval dates relative to the start of the PHE, and we expect key informant interviews will further clarify how the actual start of demonstration activities changed relative to the PHE period, which may lead to revisions in natural groupings of states. Sample size permitting, we will consider conducting meta-analysis activities within these natural groupings to ensure that we are drawing conclusions and lessons from states that were implementing activities within a similar time during the PHE.

3.6 Special Considerations for States with a SUD and SMI Demonstration

All but one of the currently approved twelve states with an approved section 1115 SMI/SED demonstration also have an approved section 1115 SUD demonstration. Alabama is the only state that does not have a section 1115 SUD demonstration. In some states, the SMI/SED and SUD demonstrations were approved concurrently, and in others, the SMI/SED demonstration was approved after the SUD demonstration. To understand how states coordinate demonstration activities, if at all, we propose to ask stakeholder interviewees about overlap and alignment between the two demonstrations, and we will review narrative text from quarterly and annual reports for relevant information that explains coordination or overlap in activities. We initially envisioned that a concurrent section 1115 SUD demonstration would be a key demonstration feature in the meta-analysis methods proposed in ***Section 3.3***. However, this will not be a useful demonstration feature given that all currently approved section 1115 SMI/SED demonstration states also have a section 1115 SUD demonstration.

Sample size permitting, we will also explore tracking monitoring metrics and impact outcomes among those with a co-occurring SMI and SUD diagnosis. If co-occurrence is

especially high, examining the co-occurring group separately from all beneficiaries with SMI may not elicit unique information, and conversely, if co-occurrence is particularly low in a state, there may not be sufficient sample size to warrant in-depth examination of this particular population.

SECTION 4.
REPORTING

Timely, useful reporting of findings and dissemination to key stakeholder audiences underlies our evaluation approach. We are engaging CMS early in the report development process to ensure that reports are developed with the agency's needs in mind and thus require fewer iterations. We are developing the outline of each report with CMS at least four months in advance of submission. Storyboarding, data visualization, and multiple rounds of technical review produce findings that are compelling, presented clearly, and accurately reflect the evidence collected. Dissemination plans stemming from these reports will be coordinated with CMS and subject to CMS approval. The content of reports will include the results available to meet CMS's information needs for each period of the evaluation. The various types of reports include:

Data Assessment Memos. Data assessment memos are meant to share with CMS our team's thinking on datasets that could support evaluation efforts and how that thinking may evolve as our team learns more about the demonstrations and CMS evaluation priorities. These memos also provide a touch point for CMS and RTI to discuss analysis plans and decisions that shape the direction of the evaluation.

- *Supplemental Secondary Data Sources: Review and Recommendations*. This memo provides a review of national datasets that may support evaluation efforts and contribute to our understanding of the prevalence of SMI/SED, mental health treatment availability, and success in meeting section 1115 SMI/SED demonstration milestones and goals. The review summarizes strengths and weaknesses of the dataset and conveys our decision as to whether the dataset is useful for the meta-evaluation. If additional national datasets are proposed or if new information becomes available (e.g., if our recommendation about whether to use the data change after we begin analyzing it), we will update this memo and our recommendations. If we do update the memo, we will also address the extent to which the dataset could be used to examine disparities in care.
- *Data Assessment: T-MSIS Analytic Files.* We will assess whether the quality and completeness of TAFs is adequate to support evaluation of SMI demonstration outcomes. We have already submitted a memo specific to the SUD demonstrations. Future data quality assessments will occur as new years of T-MSIS data become available and will incorporate a review of data elements relevant for the SMI demonstrations.
- *State Evaluation Data Memo.* We will prepare a memo that documents the states' section 1115 SMI/SED demonstration evaluation design (e.g., whether a comparison group is used, whether the comparison group is in-state or out-of-state, what outcomes will be analyzed, differences in how outcomes are specified). This information will be used to determine if it will be feasible to use interim state

evaluation results in the meta-evaluation. Final evaluation results will not be available until after this contract ends.

Rapid Cycle Reports. Rapid cycle reports will focus on emergent informational needs (e.g., an in-depth analysis of a key topic or impact analyses for policy decision support). Additional details about the rapid cycle reports are described in ***Section 4.1***.

Summative Report. Incorporating data analyzed through the three years of the evaluation, the summative report will be externally facing and document the full experience of the demonstration states, the outcomes they achieved during the demonstration period, and the patterns observed across demonstrations that can explain which policies and program components support success in meeting demonstration milestone and goals. Additional details about the summative report are found in ***Section 4.2***.

Production. Approximately four months before draft rapid cycle reports and the summative evaluation report are due, we solicit CMS's input on a draft outline that may include section headings, table shells, and proposed figures and exhibits. Report outlines are organized around key findings that address research questions to ensure all relevant sources of data are integrated into a coherent storyline. RTI's editors, graphic specialists, and data visualists work with the authors to frame their findings and develop key graphics.

Multiple stages of review for quality

- RTI senior scientific review
- Editorial review (for flow and voice)
- Copy editing
- CMS review

The process for data visualization development will begin with identifying and aligning appropriate visuals with key findings. When visualized using the proper techniques, data can be easily understood and accurately interpreted. Using the appropriate imagery, color palette, and layout, data visualizations will combine visual appeal with the data to enhance the overall impact dissemination of its findings in reports.

In preparation for each report, key technical staff, including the project director, associate project director, task leaders, and analysts, meet regularly to discuss analysis plans, findings, and interpretation of findings. Production cycles accommodate multiple rounds of senior technical reviews for each final product to ensure the content is clear, easy to understand, and all of CMS's technical directives have been fully addressed.

Timeline for Reporting. ***Exhibit 4-1*** presents the schedule of deliverables. SMI/SED evaluation efforts did not occur in the contract Base Year or Option Year 1, so these periods are not

included. We expect that the timing of data collection and reporting of various data will evolve over the course of the evaluation, so this schedule will be reviewed each year and modified as needed in coordination with CMS. If the contract is extended at any point, this timeline will be revised to align with new reporting deadlines.

Exhibit 4-1. Schedule of Deliverables

Option Year 2 (9/24/2020-9/23/2021)	Option Year 3 (9/24/21-9/23/2022)	Option Year 4 (9/24/2022-9/23/2023)
Supplemental Secondary Data Sources: Review and Recommendations	Supplemental Secondary Data Sources Update: Review and Recommendations (if needed)	Supplemental Secondary Data Sources Update: Review and Recommendations (if needed)
Evaluation Design Report	Data Assessment Update: T-MSIS Analytic Files (SMI included in SUD update)	Summative Evaluation Report: ▪ Draft report ▪ Final report due 30 days after receipt of CMS comments Briefing on the summative evaluation report
	State Evaluation Data Memo	
	4 Rapid Cycle Reports: ▪ 1 interim performance report ▪ 3 case studies 4 briefings on the rapid cycle reports	

4.1 Rapid Cycle Reports

We will prepare two types of rapid cycle report: an interim performance report and case study reports. Each rapid cycle report will contain the following:

- Purpose of the report
- Background and motivation for the report
- High-level overview of methods
- Findings
- Conclusions from findings
- Infographics and data visualizations
- Appendices with more detailed methods.

At the request of CMS, we will create a summary slide presentation for each rapid cycle report. This presentation will be used to brief CMS on the findings, methods, conclusions, and implications of each of these reports. The briefings will be conducted in person or virtually, depending on CMS's preference.

4.1.1 Interim Performance Report

One rapid cycle report produced in Option Year 3 will report on trends in service use before and after the SMI/SED demonstrations started for Medicaid enrollees with SMI in demonstration states. We will identify trends in service use from a baseline period—which will include 2016 through the calendar year prior to demonstration approval—through the first year or two of demonstration operation (which would be 2019, 2020, and 2021). Because the COVID-19 pandemic and PHE began in early 2020, this early exploration will clarify how the pandemic shifted patterns of utilization for individuals with SMI. Understanding these patterns will help shape how we address pandemic-related shifts when generating demonstration impact estimates in meta-analyses planned for the final evaluation report. Service use outcomes proposed for this analysis include:

> The interim performance rapid cycle report will focus **on trends in service use before and after the SMI/SED demonstrations started for Medicaid enrollees with SMI in demonstration states**.

- Emergency department visits
- Mental health-related admissions by select settings, including acute hospitals, psychiatric hospitals, intensive outpatient or partial hospitalization services, IMDs, and residential treatment settings
- 30-day acute care and psychiatric hospital readmissions
- Mental health-related admission discharges with a mental health-related follow-up visit within 7 and 30 days
- Total Medicaid expenditures per beneficiary.

These proposed outcomes were chosen based on their relationship to a SMI/SED demonstration goal and on our ability to calculate them using Medicaid claims data.

To explore health equity and disparities in health care, we will identify if there are any differential trends in utilization by groups defined by race, ethnicity, sex, urban/rural area of residence, and Medicaid eligibility category.

The rapid cycle report on interim performance will include infographics (multiple data visualization panels) to convey descriptive trends from the analyses and progress toward demonstration goals. Results will be organized such that a reader will clearly understand which study outcomes are related to which stated demonstration goal. The infographics will be a visually appealing product that can be shared with stakeholders and state leadership to provide an update on demonstration progress in addition to any preliminary insights on patterns in state performance.

4.1.2 Case Studies

We will prepare three rapid cycle reports in Option Year 3 that will be case studies designed to take a deep dive into targeted demonstration design and implementation topics. Reports will draw on narrative information provided by states in quarterly and annual monitoring reports, midpoint assessments if available, primary data collected from key stakeholders during interviews, and other secondary data analyses as appropriate. Report topics will be selected in consultation with CMS and will focus on emerging demonstration design and implementation issues.

Case study rapid cycle reports provide a **deep dive into select demonstration design and implementation topics** that will help inform policy directions for CMS.

Potential topics could include:

- Strategies for maintaining engagement of people in SMI/SED treatment
- Integration with social services such as supportive housing and employment services
- Enhancing access to mental health care for youth
- Delivery of crisis stabilization services under the demonstration
- Integration of physical health and SUD services into demonstration activities.

Each case study will explore how strategies to implement a particular demonstration design feature address the service needs of vulnerable populations. For example, a deep dive into strategies to maintain engagement in treatment could include interview questions ascertaining perspectives on sustaining treatment among the most vulnerable populations whose delayed care may lead to additional health concerns or involvement with the justice system. As another example, if we explored the role of crisis stabilization services within the context of health equity and health disparities, we could ask state officials to provide their perspectives on service gaps in delivery of crisis care, how it varies across particular populations, and strategies states are considering to address those gaps. Populations of interest may include children/youth with SED, individuals with SMI living in rural areas, or the individuals involved in the justice system.

In the first month of Option Year 3, RTI and CMS will discuss potential case study topics for the year and a schedule for submitting each report. Report schedules will depend on when data required for the report is expected to be available. RTI and CMS will review the planned report topics periodically during the year in the event different priorities emerge.

4.2 Summative Evaluation Report

In Option Year 4, we will submit a summative evaluation report for the section 1115 SMI/SED demonstrations. The draft summative evaluation report will be submitted in Month 5 of Option Year 4 (February 2023). The report will describe the findings from the meta-analyses

of demonstration impacts, discuss implementation experience across states, and synthesize implementation experience with impact findings to further interpret demonstration outcomes. Information from rapid cycle reports will be incorporated into the summative evaluation report as appropriate. The report will draw on the full range of data sources proposed for the meta-evaluation, including primary and secondary data. Because we will be asking states about the impact of their activities on certain groups of interest and because we propose to look at impact estimates by groups of interest to CMS, we will also incorporate findings as they relate to successful (or less successful) demonstration strategies to address observed disparities in health and health care and to promote health equity among individuals with SMI/SED.

Report preparation will begin with a series of internal analyst meetings and meetings with CMS to hone the key findings. To allow the team to review interim findings, prioritize results to be reported, and develop a cohesive story for the report, we will convene a virtual 1-day meeting for report authors. By sharing content, comments, and conclusions, this meeting will enhance our ability to provide CMS an appropriate structure, design, and focus for the report. Results from the authors' meeting will be shared with CMS to obtain the feedback needed to hone the key findings to be included in the report. Next, an outline containing section headings, table shells, and proposed data visualizations will be shared with CMS for review and comment. We will incorporate feedback on the report outline and submit a revised outline within two weeks after receiving consolidated comments from CMS.

The summative evaluation report will:

- **Detail findings and conclusions** regarding implementation practices and demonstration milestone and goal achievement across states.
- **Describe which policies, program components, and contextual factors explain more or less success in meeting demonstration goals** and affecting change in care for individuals with SMI/SED, including subpopulations of interest.
- **Highlight state strategies to address disparities in health care** or promote health equity.

To ensure that report findings are practical and valuable for stakeholders, nontechnical language will be used throughout the report; key findings will be presented using visual displays, infographics, or simplified tables; key findings and takeaways will be included at the start of chapters and sections; and each section will close with a discussion of findings. The report will also include an executive summary and a closing chapter that synthesizes results across report chapters. Appendices will be used to convey comprehensive technical detail on methods and to share more detailed tables or text supporting simplified displays of findings that will be included in the main body of the report. Finally, the

summative evaluation report will be accompanied by a concise stand-alone document of key findings that can be distributed to federal and state stakeholders.

We will submit a draft summative evaluation report to CMS for review and comment. At CMS's discretion, states and state evaluators may also review the reports. Within 30 days of receipt of CMS and other stakeholder comments, our team will provide a final summative evaluation report. As necessary, additional rounds of review from CMS and revisions by RTI will be made. Upon approval of the final summative evaluation report, we will develop a summary slide presentation of key findings for a briefing presentation to CMS. At CMS's request, we will also develop and participate in up to three webinars to disseminate the findings. These presentations might be made to states, Centers for Medicare and CHIP Services (CMCS) staff, or CMCS collaborators like the Center for Medicare & Medicaid Innovation or the Office of the Assistant Secretary for Planning and Evaluation.

REFERENCES

Bazeley, P., & Richards, L. (2000). *The NVivo Qualitative Project Book*. Sage.

CMS. (2021a). *Fact sheet: Medicaid & CHIP and the COVID-19 public health emergency.* https://www.cms.gov/newsroom/fact-sheets/fact-sheet-medicaid-chip-and-covid-19-public-health-emergency

CMS. (2021b). *Medicaid Health Homes: An Overview.* https://www.medicaid.gov/state-resource-center/medicaid-state-technical-assistance/health-home-information-resource-center/downloads/hh-overview-fact-sheet.pdf

Goodrick, D. (2014). *Comparative Case Studies*. (Methodological Briefs - Impact Evaluation, No. 9). UNICEF. https://www.unicef-irc.org/publications/754-comparative-case-studies-methodological-briefs-impact-evaluation-no-9.html

MACPAC. (2019). *Report to Congress on Oversight of Institutions for Mental Diseases*. https://www.macpac.gov/wp-content/uploads/2020/01/Report-to-Congress-on-Oversight-of-Institutions-for-Mental-Diseases-December-2019.pdf

MACPAC. (n.d.). *Behavioral health services covered under HCBS waivers and 1915(i) SPAs.* https://www.macpac.gov/subtopic/behavioral-health-services-covered-under-hcbs-waivers-and-spas/

Musumeci, M., Chidambaram , P., & Watts, M. O. (2019). *Medicaid's money follows the person program: State progress and uncertainty pending federal funding reauthorization.* https://www.kff.org/medicaid/issue-brief/medicaids-money-follows-the-person-program-state-progress-and-uncertainty-pending-federal-funding-reauthorization/

Richards, L. (2009). *Using NVIVO in Qualitative Research*. Sage.

Schneider, C. Q., & Wagemann, C. (2012). *Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis*. Cambridge University Press.

Sorensen, A. V. (2008). Media review: Use of QSR NVivo 7 qualitative analysis software for mixed methods research. *Journal of Mixed Methods Research, 2*(1).

APPENDIX A:
SELECT STATE EVALUATION PLANS: ABSTRACTION OF KEY INFORMATION

District of Columbia

Population Characteristics	
Definition of SMI	Age 21+ at start of measurement period and has either: 1) An ICD-10-CM diagnosis from the list below* during the measurement period 2) A claim with a provider type of "D05" (residential treatment center) during the measurement period, or 3) A claim with a provider type of "D02" or "D03" (public and private psychiatric hospitals) that meets the criteria of the inpatient stay HEDIS value set during the measurement period.
Definition of SED	Age <21 at start of measurement period and has either: 1) An ICD-10-CM diagnosis from the list below* during the measurement period, or 2) A claim with a provider type of "D05" (residential treatment center) during the measurement period, or 3) A claim with a provider type of "D02" or "D03" (public and private psychiatric hospitals) that meets the criteria of the inpatient stay HEDIS value set during the measurement period.
Study Population	Full-benefit Medicaid beneficiaries
Comparison group	*No comparison group will be used*

*eligible ICD10 codes: Schizophrenia: F20.0 F20.1 F20.2 F20.5 F20.81 F20.89 F20.9 F22 F25.9 F29 Mood disorders: F30.10 F30.11 F30.12 F30.13 F30.2 F30.3 F30.4 F30.8 F31.10 F31.11 F31.12 F31.13 F31.2 F31.30 F31.31 F31.32 F31.4 F31.5 F31.60 F31.61 F31.62 F31.63 F31.64 F31.73 F31.74 F31.75 F31.76 F31.77 F31.78 F31.81 F31.9 F32.0 F32.1 F32.2 F32.3 F32.4 F32.5 F32.8 F32.9 F33.0 F33.1 F33.2 F33.3 F33.41 F33.42 F33.9 F34.1 F34.8 F39; Post-traumatic stress disorder: F43.10 F43.12

Analytic Methods	
Quantitative	"The quantitative analysis will include descriptive statistics and an impact analysis using an interrupted time series (ITS) design. We will conduct descriptive subgroup analysis by stratifying the data by beneficiary characteristics, treatment setting, and service type. Descriptive statistics will include frequencies, means, and distributions of relevant metrics. ITS is the CMS-preferred methodology for impact analysis when there is no appropriate comparison group as is the case with this Demonstration. We will conduct the ITS analyses for the target population overall, as defined by each research question. In addition, we may conduct ITS analyses by treatment setting, service type, FFS and Managed Care, and dual status for selected measures, depending on sample sizes and relevance for the evaluation."
Qualitative	"We will conduct key informant interviews, site visits with providers, a beneficiary survey, and a document review to gather primary data that characterizes the interventions the District will implement to achieve the Demonstration's goals. This data will also yield insights into providers' and beneficiaries' awareness and perspectives of systems changes enacted through the Demonstration. We will employ thematic coding and triangulation to analyze the data qualitatively."

Indiana

Population Characteristics	
Definition of SMI	Identified using four diagnosis codes in the primary or secondary diagnosis position: (F20.xx (Schizophrenia and sub codes up to 2 places), F25.xx (Schizoaffective Disorder and sub codes up to two places), F31.xx (Bipolar and all sub codes up to 2 places), F33.xx (Major depression Recurrent and all sub codes up to two places)).
Definition of SED	Excluding the SED population from the evaluation
Study Population	Full benefit Medicaid beneficiaries ages 21-64
Comparison group	Study population but in the baseline period (2018-2019)

Analytic Methods	
Quantitative	"Due to the limited period of this demonstration (1 year, Jan-Dec 2020) only descriptive statistics (e.g., total, average, proportion) will be calculated (across time where necessary) as well as observational inference on trends in outcomes of interest. No inferential statistics including case-mix adjusted estimates will be developed for this evaluation."
Qualitative	"Qualitative data collected through key informant interviews will be analyzed using thematic analysis. Will support an understanding of stakeholders' perspectives related to context, implementation, and outcomes, and will identify contextual factors that help to explain outcomes."

Vermont

Population Characteristics	
Definition of SMI	Adult Medicaid members between the ages of 21 and 64 who received IMD services during the measurement period) will serve as the study population.
Definition of SED	Excluding the SED population from the evaluation
Study Population	For purposes of assessing the IMD demonstration, adult Medicaid members between the ages of 21 and 64 who received IMD services during the measurement period) will serve as the study population.
Comparison group	Study population but in the baseline period (calendar year 2019)

Analytic Methods	
Quantitative	Descriptive/inferential statistics; longitudinal design with analysis methods including logistic regression, ANOVA, and propensity score matching with T-test, as appropriate
Qualitative	No qualitative analyses were reported for the SMI study group