Medicaid 1115 Demonstration Evaluation Design Plan

Managed Long-Term Services and Supports

Design Supplement: Final Outcomes Evaluation
January 2019

Andrea Wysocki
Su Liu
Jenna Libersky
Madeline Pearse
Jonathan Gellar

This page has been left blank for double-sided copying.

CONTENTS

TABLES

This page has been left blank for double-sided copying.

I. INTRODUCTION AND PURPOSE

States are increasingly turning to managed care delivery systems, rather than fee-for-service (FFS), to provide long-term services and supports (LTSS) to Medicaid beneficiaries who are older adults or have disabilities. As of August 2018, 23 states operated 34 Medicaid managed LTSS programs (MLTSS),[1] a significant increase from the 8 states that did so in 2004 (Libersky et al. 2018; Saucier et al. 2012). MLTSS programs have the potential to provide less costly, person-centered home and community-based alternatives to institutional care, improve care quality and coordination, increase quality of life, and reduce the use of unnecessary hospital and institutional services. However, if managed care plans restrict access to services or do not assure the quality and coordination of services, MLTSS could have adverse effects on health and long-term care outcomes.

As states increasingly deliver LTSS through managed care models, it is important to understand how costs and beneficiary outcomes for MLTSS enrollees differ from those receiving LTSS through traditional FFS delivery systems. Although many states adopt MLTSS programs to control per-user spending, enhance access to home and community-based services (HCBS), and improve the quality of care, evidence on how well MLTSS programs achieve these goals has been mixed. The Centers for Medicare & Medicaid Services (CMS) commissioned Mathematica Policy Research to evaluate the performance of recent MLTSS programs[2] to examine how per-user Medicaid MLTSS spending changes over time and how MLTSS programs compare to FFS on use of specific services, access to such services, quality of care, beneficiary experience, and quality of life. Mathematica is conducting the evaluation in two rounds: an interim outcomes evaluation report was published in January 2018, and a final evaluation will be completed in 2020.

The interim evaluation report presented preliminary findings for some of these outcomes of interest (Libersky et al. 2018):

- To examine changes in per-user Medicaid MLTSS spending over time, we presented descriptive trends in annual state-level total Medicaid MLTSS and per-user Medicaid spending across all MLTSS states. We found that from 2012 to 2015, Medicaid MLTSS per-user expenditures increased by 28 percent among states that could report them.
- To examine differences in LTSS and hospital use between MLTSS and FFS systems, we compared MLTSS enrollees in two state programs—New York's Managed Long Term Care (MLTC) program and Tennessee's CHOICES program—to a similar group of people receiving LTSS under FFS. In New York and Tennessee, findings on MLTSS' ability to rebalance care from institutional settings toward home and community-based settings were mixed. In New York, the probability of using any institutional care was lower after

[1] These counts of MLTSS programs do not include programs provided under the Financial Alignment Initiative (FAI) for Medicare-Medicaid dual enrollees. We exclude the FAI programs throughout this entire report.

[2] MLTSS provided under the FAI for Medicare-Medicaid dual enrollees is being evaluated through a separate contract, which will provide additional findings about costs and beneficiary outcomes for Medicare-Medicaid dual enrollees in integrated MLTSS programs.

enrollment in MLTC, and in most instances the use of HCBS and personal care was higher relative to the FFS comparison group. In Tennessee, the probability of using personal care was higher for those who enrolled in CHOICES, but the likelihood of any use of HCBS was higher only for Medicaid-only beneficiaries and lower for dually eligible enrollees; and changes in institutional care were insignificant compared to matched FFS beneficiaries. Hospital use declined among MLTC enrollees in New York and increased among CHOICES enrollees in Tennessee. The results in New York and Tennessee were largely driven by dually eligible enrollees where a subset of this population had incomplete data.

In the final outcomes evaluation, which will focus on outcomes during the period from 2010 to 2017, we apply lessons learned from the interim outcomes evaluation and take advantage of new data; depending on its quality, the data may allow us to evaluate MLTSS program performance in more states, using more measures, and for more recent time periods.

This design report presents a framework for the final evaluation that builds on our previous MLTSS evaluation design proposals (Irvin et al. 2015; Libersky et al. 2017). First, we summarize the MLTSS program features across states (Section II). We then list the research questions for this evaluation (Section III). Next, we describe the evaluation's outcome measures and data sources (Section IV), including the steps for the data quality assessments, and proposed methods for the evaluation across different groups of states (Section V). We conclude with limitations to the evaluation (Section VI).

While Mathematica is conducting the final outcomes evaluation, our partners at IBM Watson Health will produce two companion briefs—"rapid-cycle reports"—based on semiannual tracking and analyses of demonstration implementation and progress. Findings from the rapid-cycle reports will help us interpret the findings of the outcomes evaluation, and potentially provide supplementary information that we cannot evaluate using administrative data.

II. MLTSS PROGRAM FEATURES ACROSS STATES

States that implement MLTSS programs share many common goals, such as rebalancing the LTSS system towards HCBS, improving health and functional outcomes, and reducing spending growth. However, program structure varies widely along several dimensions (Table II.1). These and other variations in program design can influence outcomes related to access, cost, and quality of care, as described below; and will affect our evaluation design as well (see Section V):

- **Start date.** Some states have been operating MLTSS programs for many years, while others have recently implemented MLTSS programs. Of the 34 MLTSS programs operating as of August 2018, 19 have begun since 2010.
- **Type of enrollment.** While most programs (24 of 34) require people to enroll in managed care to receive LTSS, 10 programs allow some groups to choose to receive LTSS through managed care (referred to as voluntary opt-in) or automatically assign them to an MLTSS plan from which they can disenroll (referred to as voluntary opt-out).
- **Populations enrolled.** Most MLTSS programs cover adults age 65 and over (31 of 34), and many cover adults with physical disabilities (24 of 34) or intellectual disabilities (22 of 34). Only 14 cover children with disabilities. All but two programs (Illinois' Integrated Care Program and Tennessee's TennCare Employment and Community First CHOICES) cover full-benefit Medicare-Medicaid enrollees—meaning that they qualify for full Medicaid and Medicare benefits, and Medicare is the primary payer for medical services.[3]
- **Level of LTSS need**. Although all programs admit people who qualify for institutional level of care, 12 programs also extend eligibility to those with low or no functional support needs (for example, Medicare-Medicaid eligible beneficiaries who qualify based on age and income).
- **Services covered by capitation.** Most programs (26 of 34) cover both Medicaid medical care and LTSS as part of a comprehensive benefit package for Medicaid-only enrollees;[4] the remaining eight programs provide LTSS through a limited-benefit managed care program separate from any programs that cover medical care ("carve out" LTSS programs).
- **Percent of counties covered by program.** Most programs (22 of 34) operate statewide, and four programs operate in greater than half of the counties in the state. Only eight programs operate in less than half of the counties in the state.

---

[3] Most partial-benefit dual eligible beneficiaries do not qualify for full state Medicaid benefits. Depending on household income, Medicaid pays either all or a share Medicare premiums, deductibles, and/or cost-sharing for these beneficiaries. For more information on categories of dual eligibility, see: https://www.cms.gov/Medicare-Medicaid-Coordination/Medicare-and-Medicaid-Coordination/Medicare-Medicaid-Coordination-Office/Downloads/MedicareMedicaidEnrolleeCategories_08012018.pdf

[4] Approximately three-quarters of Medicaid LTSS users are dual eligible beneficiaries whose acute care is covered by Medicare, either through traditional FFS Medicare, a Medicare Advantage (MA) plan, or a special MA plan, such as a Dual Eligible Special Needs Plan (D-SNP), or Fully Integrated Dual Eligible Special Needs Plan (FIDE-SNP). For more information on D-SNP contracts, see Verdier et al. (2016).

Our proposed framework for the final outcomes evaluation is influenced by the program features noted above as well as data availability. We describe our proposed data sources and methods in greater detail in Section IV and V, respectively. But first, we introduce the research questions and overall approach for addressing these questions in the final evaluation.

Table II.1. MLTSS program features, as of August 2018

State	Program name	Start date	Mandatory or voluntary enrollment	Populations enrolled: Children with dis-abilities	Populations enrolled: Adults with PD	Populations enrolled: Adults with I/DD	Populations enrolled: Older adults 65+	Populations enrolled: Full benefit Medicare-Medicaid enrollees	Minimum LOC needed to enroll	Services covered by capitation	Percent of counties covered by program
AZ	Arizona Long Term Care System (ALTCS)	1/1/1989	Mandatory	X	X	X	X	X	LTSS Less Than Institutional LOC	Medical & LTSS	100%
CA	Managed Medi-Cal Long-Term Supports and Services	4/1/2014	Mandatory		X	X	X	X	No LTSS Need	Medical & LTSS	14%[a]
CA	Senior Action Care Network (SCAN)	1/1/1996	Voluntary				X	X	Institutional LOC	Medical & LTSS	5%*
DE	Diamond State Health Plan-Plus (DSHP-Plus)	4/1/2012	Mandatory	X	X	X	X	X	No LTSS Need	Medical & LTSS	100%
FL	Statewide Medicaid Managed Care Long Term Care Program[b]	8/1/2013	Mandatory		X		X	X	Institutional LOC	LTSS Only	100%
HI	QUEST Integration[c]	1/1/2015	Mandatory	X	X	X	X	X	No LTSS Need	Medical & LTSS	100%
IA	Iowa Health Link	4/1/2016	Mandatory	X	X	X	X	X	No LTSS Need	Medical & LTSS	100%
ID	Medicare-Medicaid Coordinated Plan	7/1/2014	Voluntary – opt in		X	X	X	X	No LTSS Need	Medical & LTSS	50%
IL	HealthChoice	1/1/2018	Mandatory	X	X	X	X	X	No LTSS Need	Medical & LTSS	100%*
	Medicaid Integrated Care Program (ICP)	5/1/2011	Mandatory		X	X	X		No LTSS Need	Medical & LTSS	24%*
	Medicaid Managed Long Term Services and Supports	7/1/2016	Mandatory		X		X	X	Institutional LOC	LTSS Only	6%*
KS	KanCare (MLTSS Component)	1/1/2013	Mandatory	X	X	X	X	X	Institutional LOC	Medical & LTSS	100%
MA	Senior Care Options	3/1/2004	Voluntary – opt in				X	X	No LTSS Need	Medical & LTSS	79%*
MI	Managed Specialty Services and Supports Program[d]	1/1/1998	Mandatory	X[e]		X	X	X	Institutional LOC	LTSS Only	100%
	MI Choice	10/1/2013	Mandatory		X		X	X	Institutional LOC	LTSS Only[f]	100%
MN	Minnesota Senior Health Options (MSHO)	2/1/1997	Voluntary – opt in				X	X	Institutional LOC	Medical & LTSS	100%
	Minnesota Senior Care Plus (MSC+)	6/1/2005	Mandatory				X	X	Institutional LOC	Medical & LTSS	100%
NC	MH/IDD/SAS	4/1/2005	Mandatory			X	X	X	Institutional LOC	LTSS Only	100%

Table II.1 *(continued)*

				Populations enrolled							
State	**Program name**	**Start date**	**Mandatory or voluntary enrollment**	**Children with dis-abilities**	**Adults with PD**	**Adults with I/DD**	**Older adults 65+**	**Full benefit Medicare-Medicaid enrollees**	**Minimum LOC needed to enroll**	**Services covered by capitation**	**Percent of counties covered by program**
NJ	NJ FamilyCare (MLTSS Component)	7/1/2014	Mandatory		X		X	X	Institutional LOC	Medical & LTSS	100%
NM	Centennial Care (MLTSS Component)[g]	1/1/2014	Mandatory	X	X	X	X	X	Institutional LOC	Medical & LTSS	100%
NY	MLTC Partial Capitation	1/1/1998	Mandatory		X		X	X	Institutional LOC	LTSS Only	94%*
	Medicaid Advantage Plus	10/1/2007	Voluntary – opt in		X		X	X	Institutional LOC	Medical & LTSS	66%*
	FIDA/IDD	4/1/2016	Voluntary – opt in			X	X	X	Institutional LOC	Medical & LTSS	15%*
OH	MyCare Opt-out[h]	5/1/2014	Mandatory		X	X	X	X	No LTSS Need	Medical & LTSS	33%*
PA	Adult Community Autism Program	1/1/2009	Voluntary – opt in			X[i]	X	X	Institutional LOC	LTSS Only	6%
	Community HealthChoices	1/1/2018	Mandatory		X	X[j]	X	X	No LTSS Need	Medical & LTSS	21%[k]
RI	Rhody Health Options (MLTSS Component)	11/1/2013	Voluntary opt out		X	X	X	X	No LTSS Need	Medical & LTSS	100%
TN	TennCare CHOICES in Long-Term Care	3/1/2010	Mandatory	X[l]	X		X	X	LTSS Less Than Institutional LOC	Medical & LTSS	100%
	Employment and Community First CHOICES	7/1/2016	Mandatory	X[m]		X		X	LTSS Less Than Institutional LOC	Medical & LTSS	100%
TX	Texas STAR+PLUS	1/1/1998	Mandatory	X[n]	X	X	X	X	No LTSS Need	Medical & LTSS	100%
	Texas STAR Kids	11/1/2016	Mandatory	X				X	No LTSS Need	Medical & LTSS[f]	100%
	Texas STAR Health	4/1/2008	Voluntary opt out	X					No LTSS Need	Medical & LTSS	100%
VA	Commonwealth Coordinated Care Plus[o]	8/1/2017	Mandatory	X	X	X	X	X	Institutional LOC	Medical & LTSS	100%
WI	Family Care	1/1/1999	Voluntary – opt in		X	X	X	X	LTSS Less Than Institutional LOC	LTSS Only	100%
	Family Care Partnership	1/1/1996	Voluntary – opt in		X	X	X	X	Institutional LOC	Medical & LTSS	19%*

Source: Unpublished program features data provided by IBM Watson Health, August 2018.

Note: Information is current as of August 2018. This table does not include MLTSS programs provided under the CMS Medicare-Medicaid FAI. HCBS = home and community-based services; I/DD = intellectual or developmental disabilities; ICF/IDD = Intermediate care facilities for individuals with intellectual disabilities; LOC = level of care; LTSS = long-term services and supports; MLTSS = managed long-term services and supports; NF = nursing facility; PD = physical disabilities.

*Includes the most populous counties in the state.

[a] Two of the eight counties have not yet enrolled members: Alameda and Orange (1115 demonstration approval, attachment U "CCI Enrollment Timeline by Population and County").

Table II.1 *(continued)*

[b] The Florida Long-Term Care Community Diversion Program (MLTSS program with 1915a/1915c authority that began in 1997) was phased out in 2014; from August 2013 through March 2014 the state transitioned Long-Term Care Community Diversion Program members into the current, now mandatory, program.

[c] Hawaii's QUEST Expanded Access program, or QExA, (MLTSS program with 1115 authority that began in 2009) was combined with the QUEST managed care program to cover all Medicaid managed care through one program, QUEST Integration, as of January 2015.

[d] Michigan’s Specialty Services and Supports Program covers mental health and substance use disorder services, and LTSS for all Medicaid beneficiaries with mental illnesses, substance use disorders, or developmental disabilities through county-based prepaid inpatient health plans (PIHPs). According to data collected by CMS in 2017, only 7,634 of the total 2,286,950 enrollees use LTSS. Because the program predominantly serves non-MLTSS users, we will not consider the program for inclusion in our outcomes evaluation. However, because Michigan reports some LTSS expenditures, we are including it in our expenditure analysis.

[e] Children with serious emotional disturbance (SED) and/or DD.

[f] Program includes HCBS only (NF and ICF/IDD are carved out).

[g] New Mexico's CoLTS mandatory MLTSS program (1915b/1915c authority) began in 2008. In January 2014, New Mexico consolidated the administration of CoLTS and its managed care program Salud! through a new 1115 demonstration referred to a Centennial Care. The new program covers behavioral health benefits for MLTSS enrollees, while the previous MLTSS program provided behavioral health benefits through a separate behavioral health managed care program.

[h] Ohio requires that dually eligible beneficiaries enroll in one of two service options, both referred to as MyCare: (1) a FAI demonstration that integrates Medicare and Medicaid benefits through Medicare-Medicaid plans, or (2) an MLTSS program for beneficiaries who opt out of the FAI demonstration that provides LTSS through non-integrated managed care plans. This table presents information on the MLTSS opt-out program only.

[i] Must have a diagnosis of Autism Spectrum Disorder.

[j] People who receive waiver or other services from the Office of Developmental Programs are excluded, but other dual eligible beneficiaries with I/DD are included.

[k] Phase 1 of three phases. Statewide coverage planned by January 1, 2020.

[l] Children in nursing homes only.

[m] Children with I/DD.

[n] This group is not mandatory.

[o] Virginia's Commonwealth Coordinated Care Initiative was the state's FAI program that began in 2013 (1932a/1915c authority) and phased out as a new MLTSS program was phased in; in 2017, the state began operating the current MLTSS program: Commonwealth Coordinated Care Plus. All FAI beneficiaries were transitioned to the current, now statewide, program by January 1, 2018.

This page has been left blank for double-sided copying.

III. RESEARCH QUESTIONS

The goal of this evaluation, conducted under the National 1115 Demonstration Evaluation, is to understand how LTSS-related outcomes at the program and beneficiary levels differ between managed care and FFS. Specifically, the evaluation will address five research questions:

- How does service use compare between MLTSS and FFS systems, and by MLTSS program features?
- How does the quality of care compare between MLTSS and FFS systems, and by MLTSS program features?
- How does self-reported access to care compare between MLTSS and FFS systems, and by MLTSS program features?
- How does beneficiary experience and quality of life compare between MLTSS and FFS systems, and by MLTSS program features?
- How does Medicaid MLTSS spending change over time, and vary by MLTSS program features?

This page has been left blank for double-sided copying.

IV. OUTCOME MEASURES AND DATA SOURCES

A. Outcome measures

Corresponding to each research question, we have selected outcome measures and data sources that are of high policy relevance and feasible to construct with available data (Table IV.1), pending validation of data quality as discussed further in Section IV.

- For service use, we propose to evaluate eight claims-based measures for Medicaid-only and dual eligible beneficiaries that reflect use of different types of LTSS and one claims-based measure—hospital inpatient days—for Medicaid-only beneficiaries that reflects hospital utilization.[5]
- For quality of care, we propose to evaluate one claims-based LTSS measure—minimizing institutional length of stay—and one claims-based medical care measure—potentially avoidable hospitalization rate—both for Medicaid-only beneficiaries.[6]
- For experience of and access to care, and quality of life, we propose to summarize survey findings across six domains, many of which are not easily measured through claims (Libersky 2018).
- Finally, for spending, we will analyze two measures to understand how Medicaid MLTSS spending changes over time, including Medicaid MLTSS expenditures by service category, and Medicaid MLTSS expenditures per user by state. MLTSS service categories will include Medicaid-paid short and long-stay nursing facilities; Intermediate Care Facilities for Individuals with Intellectual Disabilities (ICFs/IID); personal care; home health; HCBS under managed care authorities, such as Section 1115 demonstrations, Section 1915(b) waivers, Section 1915(a) contracts, and Section 1932(a) state plan amendments; HCBS under 1915(c) waivers; and other unspecified HCBS.

[5] Medical care for the dual eligible beneficiaries included in this evaluation is covered by a different plan than the MLTSS plan (for example, by a separate non-integrated MA plan or by Medicare FFS). This arrangement suggests that we should limit the analysis of medical care outcomes to Medicaid-only beneficiaries. First, MA enrollment increased from 10 percent in 2006 to 29 percent in 2016 among full benefit dual eligible beneficiaries (CMS Medicare Medicaid Coordination Office [MMCO] 2017). Although we could obtain Medicare FFS claims, MA encounter data are not available for most years of our analysis; therefore, we are unable to construct medical care measures for a large proportion of the dual eligible MLTSS population in MA plans for most of the time period covered in this evaluation. Second, non-integrated MLTSS plans are not liable for medical services paid for by Medicare, so they have no direct financial incentive or mechanism to control medical service utilization.

[6] The LTSS-related quality of care measure (minimizing institutional length of stay) also requires Medicare data to construct the measure for dual eligible beneficiaries. Because MA encounter data are not available for most years of our analysis and because non-integrated MLTSS plans are not incentivized to control medical spending (see above), we will limit the analysis of the minimizing institutional length of stay outcome measure to Medicaid-only beneficiaries.

Table IV.1. Outcome measures and data sources by research question

Measure	Definition	Population or level of aggregation	Data source	Analysis
How does service use compare between MLTSS and FFS systems, and by MLTSS program features?				
HCBS use	Percentage of the study population who use any HCBS in a month	Medicaid-only and dually eligible beneficiaries	MAX/AlphaMAX/ TAF	Matched comparison group regression and Bayesian metaanalysis
ILTC use	Percentage of the study population who use any ILTC in a month	Medicaid-only and dually eligible beneficiaries	MAX/AlphaMAX/ TAF	Matched comparison group regression and Bayesian metaanalysis
Round-the-clock services use[a]	Percentage of the study population who use any round-the-clock services in a month	Medicaid-only and dually eligible beneficiaries	MAX/AlphaMAX /TAF	Matched comparison group regression and Bayesian metaanalysis
Day services use[a]	Percentage of the study population who use any day services in a month	Medicaid-only and dually eligible beneficiaries	MAX/AlphaMAX/ TAF	Matched comparison group regression and Bayesian metaanalysis
Home-delivered meals use[a]	Percentage of the study population who use any home-delivered meals in a month	Medicaid-only and dually eligible beneficiaries	MAX/AlphaMAX/ TAF	Matched comparison group regression and Bayesian metaanalysis
Home-based services use[a]	Percentage of the study population who use any home-based services in a month	Medicaid-only and dually eligible beneficiaries	MAX/AlphaMAX/ TAF	Matched comparison group regression and Bayesian metaanalysis
Caregiver support services use[a]	Percentage of the study population who use any caregiver support services in a month	Medicaid-only and dually eligible beneficiaries	MAX/AlphaMAX/ TAF	Matched comparison group regression and Bayesian metaanalysis
Equipment, technology, and modifications use[a]	Percentage of the study population who use any equipment, technology, or modifications in a month	Medicaid-only and dually eligible beneficiaries	MAX/AlphaMAX/ TAF	Matched comparison group regression and Bayesian metaanalysis
Inpatient hospital days	Average number of inpatient hospital days each month for people in the study population	Medicaid-only beneficiaries	MAX/AlphaMAX/ TAF	Matched comparison group regression and Bayesian metaanalysis
How does the quality of care compare between MLTSS and FFS systems, and by MLTSS program features?				
Minimizing institutional length of stay	Percentage of the study population admitted to an institutional facility who are successfully discharged to the community (community residence for 60 or more days) within 100 days of admission	Medicaid-only beneficiaries	MAX/AlphaMAX/ TAF	Matched comparison group regression and Bayesian metaanalysis
Potentially avoidable hospitalizations	Percentage of the study population who have at least one potentially avoidable hospitalization due to an ambulatory care sensitive condition (AHRQ ACSC PQI #90) in a month	Medicaid-only beneficiaries	MAX/AlphaMAX/ TAF	Matched comparison group regression and Bayesian metaanalysis

Table III.1 *(continued)*

Measure	Definition	Population or level of aggregation	Data source	Analysis
How does self-reported access to care compare between MLTSS and FFS systems, and by MLTSS program features?				
NCI-AD measures[b]	Percentage of MLTSS and FFS-LTSS users who: • Have transportation to get to medical appointments when they need to • Have paid support staff that show up and leave when they are supposed to • Always get enough assistance with everyday activities (like preparing meals, housework, shopping, or taking their medications) or self-care (like bathing, dressing, going to the bathroom, eating, or moving around the home) when they need it • Can reach their case manager/care coordinator when they need to	All MLTSS enrollees and FFS users	NCI-AD data or summaries in national reports (NASUAD and HSRI 2017)	Descriptive trends for MLTSS and FFS states that fielded the survey (see Table IV.3) and Bayesian meta-analysis
How does beneficiary experience and quality of life compare between MLTSS and FFS systems, and by MLTSS program features?				
NCI-AD measures[b]	Percentage of MLTSS and FFS-LTSS users who: • Have paid support staff that do things the way they want them done • Reported feeling comfortable and supported enough to go home after being discharged from a hospital or rehabilitation facility • Reported someone followed up with them after discharge from a hospital or rehabilitation facility • Receive services that meet all of their needs and goals • Like where they are living • Feel safe at home and/or around their paid support staff • Feel in control of their life	All MLTSS enrollees and FFS users	NCI-AD data or summaries in national reports (NASUAD and HSRI 2017)	Descriptive trends for MLTSS and FFS states that fielded the survey (see Table IV.3) and Bayesian meta-analysis
How does MLTSS spending change over time, and vary by MLTSS program features?				
MLTSS expenditures by category of service	Total MLTSS expenditures 2013-2017 for the following categories of service: • ILTC (nursing facility, ICF/IID) • HCBS (personal care, home health, HCBS under managed care authorities, and HCBS under 1915(c) waivers)	All MLTSS enrollees	CMS-64/LTSS Expenditure reports	Descriptive trends for all MLTSS states and Bayesian meta-analysis

[a] The final list of HCBS that will be included as outcome measures will depend on our data quality reviews and sample sizes. The list of HCBS included in this table represent the services we will explore, but it is possible that we will be unable to include some of these categories of services in our final list of outcome measures.

[b] The proposed measures were selected to inform questions related to quality of care, access to care, and experience/quality of life that are not easily measured through encounter or claims-based measures; the final list of measures presented in the final report is subject to change.

ACSC = ambulatory care sensitive condition; AHRQ = Agency for Healthcare Research and Quality; FFS = fee-for-service; HCBS = home and community-based services; HSRI = Human Services Research Institute; ICF/IID = Intermediate Care Facilities for Individuals with Intellectual Disabilities; ILTC = institutional long-term care; MAX = Medicaid Analytic eXtract; MLTSS = managed long-term services and supports; NASUAD = National Association of States United for Aging and Disabilities; NCI-AD = National Core Indicators – Aging and Disabilities; PQI = Prevention Quality Indicators; TAF = T-MSIS Analytic File.

B. Medicaid Administrative Data

For outcome measures related to LTSS and hospital service utilization, we will use a combination of national Medicaid administrative data sources. Because we are conducting a multi-state evaluation, we will rely on national Medicaid administrative data from Medicaid Analytic eXtract (MAX), Alpha-MAX, and the Transformed Medicaid Statistical Information System (T-MSIS) Analytic File (TAF) from 2010-2017.[7,8] Our study period (2010-2017) covers the time when the retired Medicaid Statistical Information System (MSIS) was replaced by TMSIS as the national, uniform, and comprehensive data collection system for Medicaid and Children's Health Insurance Program (CHIP). For periods before a state's transition, we will use MAX, or the early version of MAX known as Alpha-MAX that does not require as many quarters of run-out for claims adjustments. For periods after a state's transition, we will use TAF. MAX and Alpha-MAX are both research versions of state MSIS submissions; TAF is a research version of state T-MSIS submissions.[9,10]

The exact data source to be used will vary by state and by year, depending on data availability and quality, as well as when each state transitioned their data systems. Table IV.2 provides a detailed look at the availability of each data source for our study period, by state. States with and without MLTSS programs are included in the table, since the latter are likely to be considered for constructing comparison groups for evaluating specific MLTSS programs. This table reflects information as of September 6, 2018 and assumes that our evaluation will take advantage of all data available by December 2018, including 2014 and 2017 TAF which are currently in production and expected to be available this calendar year.[10] The large majority of states should have data available throughout our entire study period, albeit a few states may be missing some run-out months for claims adjustment in Alpha-MAX. Following Table IV.2, we propose our approach to selecting the states to include in the evaluation based on data availability, completeness, and quality. Our assessment focuses on encounter data documenting the services used under capitated managed care, which is critical for this MLTSS evaluation.

[7] Beginning in 2010, MAX began applying the HCBS taxonomy to FFS HCBS claims submitted under 1915(c) waivers. For the final evaluation, we will replicate the methodology that MAX uses to apply the HCBS taxonomy to encounter claims in all years, and to FFS waiver claims in 2009 for programs that started in 2010 in order to have pre-period data for the 2010 programs.

[8] CMS required that states transition from reporting Medicaid administrative data from MSIS (the source of MAX and Alpha-MAX data) to T-MSIS (the source of TAF data); however, actual dates of transition vary by state. See Table IV.2 for more information on data availability by source.

[9] CMS develops MAX data as a more research-friendly version of MSIS files and TAF as a more research-friendly version of T-MSIS files.

[10] 2016 TAF data are already available.

Table IV.2. Likely Medicaid administrative data source to be used for evaluating claim-based MLTSS outcome measures, by state and year

State	2010	2011	2012	2013	2014	2015	2016	2017
States with MLTSS programs								
AZ	MAX	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF
CA	MAX	MAX	MAX	MAX	MAX	Alpha-MAX[b]	TAF	TAF
DE	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF	TAF
FL	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF	TAF
HI	MAX	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF
IA	MAX	MAX	MAX	MAX	MAX	Alpha-MAX[b]	TAF	TAF
ID	MAX	MAX	MAX	MAX	MAX	Alpha-MAX[b]	TAF	TAF
IL	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF	TAF
KS	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF	TAF
MA	MAX	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF
MI	MAX	MAX	MAX	MAX	MAX	Alpha-MAX[b]	TAF	TAF
MN	MAX	MAX	MAX	MAX	MAX	Alpha-MAX[b]	TAF	TAF
NC	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF	TAF
NJ	MAX	MAX	MAX	MAX	MAX	Alpha-MAX[b]	TAF	TAF
NM	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF	TAF
NY	MAX	MAX	MAX	MAX	Alpha-MAX[a]	Alpha-MAX[b]	TAF	TAF
OH	MAX	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF
PA	MAX	MAX	MAX	MAX	MAX	Alpha-MAX[b]	TAF	TAF
RI	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF	TAF
TN	MAX	MAX	MAX	MAX	MAX	Alpha-MAX[b]	TAF	TAF
TX	MAX	MAX	MAX	Alpha-MAX[a]	Alpha-MAX[a]	TAF	TAF	TAF
VA	MAX	MAX	MAX	Alpha-MAX[a]	Alpha-MAX[a]	TAF	TAF	TAF
WI	MAX	MAX	MAX	Alpha-MAX[b]	TAF	TAF	TAF	TAF
States without MLTSS programs								
AL	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF	TAF
AK	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF	TAF
AR	MAX	MAX	MAX	MAX	Alpha-MAX[a]	Alpha-MAX[b]	TAF	TAF
CO	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF	TAF
CT	MAX	MAX	MAX	MAX	Alpha-MAX[a]	Alpha-MAX[b]	TAF	TAF
DC	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF	TAF
GA	MAX	MAX	MAX	MAX	MAX	Alpha-MAX[b]	TAF	TAF
IN	MAX	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF
KY	MAX	MAX	MAX	Alpha-MAX[a]	Alpha-MAX[a]	TAF	TAF	TAF
LA	MAX	MAX	MAX	MAX	MAX	Alpha-MAX[b]	TAF	TAF
ME	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF	TAF
MD	MAX	MAX	MAX	Alpha-MAX[a]	Alpha-MAX[a]	TAF	TAF	TAF
MS	MAX	MAX	MAX	MAX	MAX	Alpha-MAX[b]	TAF	TAF
MO	MAX	MAX	MAX	MAX	MAX	Alpha-MAX[b]	TAF	TAF
MT	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF	TAF
NE	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF	TAF
NV	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF	TAF
NH	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF	TAF
ND	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF	TAF
OK	MAX	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF
OR	MAX	MAX	MAX	MAX	Alpha-MAX[a]	Alpha-MAX[b]	TAF	TAF
SC	MAX	MAX	MAX	Alpha-MAX[a]	Alpha-MAX[a]	TAF	TAF	TAF
SD	MAX	MAX	MAX	MAX	MAX	Alpha-MAX[b]	TAF	TAF
UT	MAX	MAX	MAX	MAX	MAX	Alpha-MAX[b]	TAF	TAF
VT	MAX	MAX	MAX	MAX	MAX	Alpha-MAX[b]	TAF	TAF
WA	MAX	MAX	MAX	MAX	Alpha-MAX[a]	TAF	TAF	TAF
WV	MAX	MAX	MAX	MAX	MAX	Alpha-MAX[b]	TAF	TAF
WY	MAX	MAX	MAX	MAX	MAX	Alpha-MAX[b]	TAF	TAF

Note: Information updated as of September 6, 2018. When both MAX/Alpha-MAX and TAF are available for the same year, we will use MAX/Alpha-MAX.

Table IV.2 *(continued)*

[a] T-MSIS files were used as inputs for part or all of the production year.

[b] Fewer than seven quarters of MSIS data were used to construct the Alpha-MAX files, in which case, we will use caution when analyzing the data, recognizing some adjustment claims may be missing.

MAX = Medicaid Analytic eXtract; TAF = T-MSIS Analytic File.

To select states to evaluate, we will assess each state's Medicaid data quality by source and study year. Given TAF is the newest Medicaid administrative data source and is still under development, we will focus on its completeness and anomalies, by state and by year, using the most recent data run available as of December 2018. Our TAF data quality assessment will borrow from the ongoing data quality monitoring work conducted by Mathematica under CMS's Medicaid and CHIP Business Information Solution (MACBIS) initiative. Certain data quality issues, such as missing populations, missing values for key variables, incorrect coding schemes, or low claims volume, will undoubtedly affect our ability to include some states in the evaluation.

Because service use for LTSS covered under managed care is contained in encounter records, which historically have been poorly reported by states, our data quality review will pay special attention to encounter data quality in both MAX/Alpha-MAX and TAF. For each state and time period, we will use an approach similar to the MAX/Alpha-MAX data quality analysis conducted for the interim evaluation report (Libersky et al. 2017). Our approach is summarized below:

1. Identify states and time periods that have more than one encounter record for LTSS (HCBS and institutional services) reported in each year during the study period. Exclude states without any LTSS encounter records from the claims-based outcome evaluation.
2. Among remaining states, check the number of beneficiaries identified as MLTSS enrollees against external benchmarks (for example, Medicaid managed care enrollment data reported on data.Medicaid.gov and in Saucier et al. 2012). Exclude states with reported MLTSS enrollment in MAX/Alpha-MAX/TAF that differs by a large margin from reported MLTSS enrollment in other sources (for example, states that have a difference of more than 25 percent).
3. Among remaining states where individuals who do not meet an institutional level of care can enroll (specifically, people who are dually eligible or who require a less than institutional level of care can enroll), examine the quality of level of care status reporting in TAF. Excludes states in which we cannot identify the correct proportion expected to meet an institutional level of care.
4. Among remaining states, for the identified MLTSS enrollees in each state, examine the percentage of enrollees who had at least one HCBS or institutional care encounter record in each study year. We would expect the percentage of MLTSS enrollees who are eligible based on an institutional level of care with at least one LTSS encounter would be nearly 100 percent. We will exclude states where enrollees have too few LTSS encounters (for example, less than 50 percent of all MLTSS program enrollees) from further analyses.
5. Among remaining states, examine the quality of the data fields on HCBS and institutional care encounter records required to construct the claims-based outcome measures identified above. For example, do institutional care encounter records have a complete service begin date and end date? Do HCBS encounter records, grouped according to the HCBS taxonomy,

have complete information related to service provision, such as place of service code or procedure code? When applicable, we will compare the measures of encounter data quality to similar measures that have been constructed for FFS claims. We will exclude states with encounter data in which key data fields required to measure the majority of the claims-based outcome measures is of poor quality.

Any MLTSS states that pass the above quality checks will be included in our final outcomes evaluation. We will determine the most appropriate states without MLTSS to include on a program-by-program basis, described further in Section V. For states in which data problems exist only in some years, we will evaluate whether we can include the state in the analysis within the existing evaluation design framework for years where data quality is acceptable.

C. Survey data on beneficiary access, experience of care, and quality of life

To evaluate beneficiary access, experience of care, and quality of life, we will use beneficiary-reported outcomes from the National Core Indicators – Aging and Disabilities (NCIAD).[11] NCI-AD is a voluntary survey available to state Medicaid, aging, and disability agencies to assess the quality of life and outcomes of seniors and adults with physical disabilities who are accessing publicly funded services, including MLTSS. NCI-AD applies to any publicly funded LTSS program, including those that cover nursing facilities (National Association of States United for Aging and Disabilities [NASUAD] 2018). The survey collects information on key facets of LTSS, such as service and care coordination, community participation, choice and decision making, employment, rights and respect, health care and safety (NASUAD and Human Services Research Institute [HSRI] 2017).[12] NCI-AD was first fielded between June 1, 2015 and May 31, 2016 among seven MLTSS states and four FFS-only states (Table IV.3). Only two MLTSS states (New Jersey and Tennessee) and four FFS-only states also fielded the survey the following year (NASUAD 2018). NASUAD and HSRI produce comprehensive reports of survey results by state and across states, and make them available on the NASUAD website.

We are exploring the possibility of obtaining beneficiary-level NCI-AD data from NASUAD and HSRI to compare access and beneficiary experience between MLTSS and FFS beneficiaries in each state, while controlling for beneficiary characteristics across these two groups. Obtaining beneficiary-level data will allow us to exclude MLTSS enrollees who reside in institutions from our sample. This exclusion will make the MLTSS sample more comparable to the beneficiaries in FFS-only states (where institutional residents are excluded from the HCBS waiver programs we would use as a comparison group). In addition, the beneficiary-level data will allow us to determine the sample size by LTSS program for each question we include in our

[11] More information on NCI-AD is available at https://nci-ad.org/about/.

[12] Another survey, the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Home and Community Based Services Survey (HCBS CAHPS), is also available for states with HCBS programs to evaluate experience of care among adults with different disabilities, including frail elderly, individuals with physical disabilities, persons with developmental or intellectual disabilities, those with acquired brain injury, and persons with severe mental illness. Between October 2013 and March 2015, three states with MLTSS and seven states with only FFS LTSS field-tested the survey. Only one MLTSS state (Arizona) and five FFS-only states also conducted the survey in 2017. Because Arizona is the only MLTSS state that conducted the survey at two time points, our ability to use the HCBS CAHPS data for this evaluation is limited. As more states and years become available, HCBS CAHPS has the potential to be an important data source for future MLTSS evaluation work.

analysis, which will improve our estimates in the meta-analysis. In the event that we are unable to obtain the beneficiary-level data, we will use the aggregate survey findings published by NASUAD and HSRI to summarize state-level descriptive trends in our report.

Using the best data we are able to obtain, we will also conduct a meta-analysis to summarize trends across MLTSS and FFS states (see Section V.B). Such trends may provide context for our claims-based outcome findings and potentially identify future research questions or study approaches that will be possible when survey data from additional states and years are available.

Table IV.3. NCI-AD availability and number of respondents among MLTSS and FFS-only states, by survey year

	2015-2016		2016-2017	
Sample size	MLTSS	FFS[a]	MLTSS	FFS[a]
MLTSS states[a]				
Delaware	314	92	-	-
Kansas[b]	-	197	-	209
Minnesota[b]	1,224	2,140	-	388
New Jersey	415	104	669	-
Ohio[b, c]	-	256	307	918
Tennessee	923	-	852	-
Texas	1,457	-	-	-
FFS-only states[a]				
Colorado	N/A	316	N/A	316
Georgia	N/A	331	N/A	-
Indiana	N/A	696	N/A	708
Maine	N/A	437	N/A	462
Mississippi	N/A	936	N/A	965
North Carolina	N/A	611	N/A	-
Nevada	N/A	-	N/A	396
Oregon	N/A	-	N/A	517

Sources: NASUAD and HSRI 2017; NASUAD and HSRI 2018a-2018l.

Notes: N/A = The state does not operate MLTSS.

[a] All states with MLTSS programs also provide some LTSS through FFS for certain populations. For this reason, we categorized states as MLTSS (with or without FFS LTSS) and FFS-only.

[b] Though Kansas, Minnesota, and Ohio operated MLTSS programs during the survey years, results of the NCI-AD survey were limited to FFS-LTSS users in one or more survey years.

[c] The NCI-AD survey includes the FAI demonstration population as well as the MyCare Opt-out (that is, the MLTSS program for dual beneficiaries).

FFS = fee-for-service; MLTSS = managed long-term services and supports; NCI-AD = National Core Indicators – Aging and Disabilities.

D. Data on MLTSS spending

Similar to the interim outcomes evaluation, the final outcomes evaluation will report descriptive trends in spending across all states that have implemented MLTSS at any point prior to 2017. It will present three indicators for each state: (1) MLTSS expenditures by service category; (2) estimated MLTSS expenditures per enrollee; and (3) the number of MLTSS enrollees (or in some states, the number of MLTSS enrollees who use LTSS), which will serve as a denominator for per enrollee spending.[13] The final evaluation will build on the information reported in the interim outcomes evaluation by adding data from 2016 and 2017; it will also correct any data errors identified in previous years.

We will calculate indicators on spending using two data sources:

- **LTSS expenditure reports.** On a quarterly basis, states report their Medicaid expenditures to CMS with Form CMS-64, which is used to determine the amount of Federal Financial Participation (FFP) they receive. Since 2008, IBM Watson Health has collected details on LTSS-specific expenditures under managed care to supplement the CMS-64 data. IBM summarizes this data each year in LTSS annual expenditure reports. We will use this detailed data to calculate MLTSS annual spending measures and will discuss any relevant limitations with IBM.
- **Medicaid Managed Care Data Collection System (MMCDCS).** Since 2013, CMS (via Mathematica Policy Research) has collected state-level counts of MLTSS enrollees who use any LTSS during the year, as well as plan-level enrollment in managed care programs including MLTSS programs. We will use this data on state-level counts of MLTSS and/or LTSS users, depending on the state, as the denominator for measures that are defined as rates per enrollee or user, as appropriate. Where necessary, we will supplement MMCDCS data with more current information reported on state websites.

[13] Information on member-months of enrollment is not available for all MLTSS states; therefore, we will use a point-in-time count of MLTSS enrollees. Point-in-time counts will over-estimate enrollment relative to using member-months of enrollment.

This page has been left blank for double-sided copying.

V. PROPOSED ANALYTIC METHODS

To address each research question using the corresponding outcome measures and data sources described above, we will employ different methods to evaluate impacts of the MLTSS programs, employing the most rigorous methods feasible. For claims-based outcome measures of service use and quality of care (research questions 1 and 2), we will use rigorous propensity score matching to ensure the MLTSS and comparison FFS groups are comparable, followed by difference-in-differences (DID) or repeated cross sectional (RCS) regression models to estimate impacts. The choice between DID and RCS models will depend on each MLTSS program's design features as well as data availability (see Section V.A).

We will use descriptive and trend analysis to examine access, beneficiary experience, and quality of life (research questions 3 and 4) for a limited number of MLTSS and FFS states and to examine expenditure measures (research question 5) for all MLTSS states, respectively (Sections V.B and V.C). In addition, to examine the relationship between program features and key outcomes, we will perform a cross-program/state meta-analysis for each outcome measure.

A. Methods for examining service use and quality of care (research questions 1 and 2)

To understand the impacts of MLTSS programs on service use and quality of care, we would like to compare outcomes among MLTSS enrollees to the outcomes that would have been observed had they not participated in the program. Of course, the latter scenario (known as the *counterfactual*) is impossible to observe, so we will estimate counterfactual outcomes using an appropriate comparison group of FFS LTSS users. For each outcome, we define the program impact to be the difference in expected outcomes under MLTSS and FFS, averaged over all MLTSS enrollees.

The variation in MLTSS program features and data availability across states prohibits the use of a single impact analysis that pools together MLTSS programs from multiple states. Instead, we must conduct a separate impact evaluation for each MLTSS program that passes our feasibility and data quality assessment. There are multiple reasons for this. First, each MLTSS program is implemented differently—for example, for different LTSS target populations—which could result in different program impacts on service use and quality of care. Second, the LTSS environment varies substantially across states. For instance, some states made significant progress in rebalancing their LTSS system in the early 2000s, while other states have made progress in rebalancing in recent years. Third, data availability and quality precludes certain evaluation approaches in certain states. Our separate evaluations for each program will result in program-level estimates of the impact of MLTSS on service use and quality of care, generated by distinct matched comparison groups and regression models. Our framework for the comparison group selection and model specification is described in detail below.

Within each program, we will produce estimates for each year and, to the extent possible, estimates for key subgroups. At minimum, we will produce separate estimates for dual eligible beneficiaries and Medicaid-only beneficiaries. If feasible, we will also consider producing estimates for other subgroups, such as for specific LTSS subpopulations defined by age (for example, adults under age 65and adults age 65 and older).

1. Choosing a design approach for each program evaluated

a. Key design considerations

In this section, we discuss two key considerations that will affect the design we use to evaluate each MLTSS program.

Over-arching design: DID or RCS. For each program that meets our data quality and availability criteria, we will employ one of two over-arching designs: either a DID design, or a RCS design. In general, the DID design is preferred. The main difference between the two designs is that the DID design removes unobserved differences between the populations receiving MLTSS and the comparison group that existed prior to program implementation from our impact estimates. This feature gives our impact estimates an added level of robustness. Nonetheless, there are two key requirements that need to be met in order to implement a DID design:

- First, we need to clearly define (in claims) a group of individuals who would receive MLTSS and a comparison group before and after the start of the program. The MLTSS group must be defined from a region where an MLTSS program will eventually be implemented, and the comparison group must be defined from a region that will never have an MLTSS program.
- Second, we must be able to observe impacts for the MLTSS and comparison groups both before and after the start of the program. For example, we cannot measure outcomes if we do not have access to MAX/TAF data in the years before the program was implemented.

If either of these criteria are not met, we will implement an RCS design. Though we are unable to remove pre-existing differences between the MLTSS and FFS groups with an RCS design, it still produces valid impact estimates as long as the two groups are well-balanced on factors that are not directly impacted by participation in the program, such as demographic factors and chronic conditions. Both DID and RCS designs control for secular trends over time that affect both MLTSS and FFS beneficiaries equally. We provide more details on both DID and RCS designs, including regression model specifications, in Section V.A.3.

Comparison group location: in-state or out-of-state. Another key design consideration is whether the comparison group is selected within the same state as the MLTSS program, or whether it is selected from another state or states. The use of an in-state comparison group would eliminate a potential source of bias emerging from state-specific differences in LTSS implementation and the surrounding Medicaid environment. However, this option is only available if the state has a sufficient number of FFS LTSS beneficiaries in the program state after the MLTSS program starts, and if these FFS LTSS beneficiaries are comparable to the MLTSS enrollees. In particular, an in-state comparison group will not be possible if the program is mandatory state-wide for all LTSS-eligible beneficiaries.

b. Program features that affect the design

There are three key program features that affect the design choices discussed above:

1. Program start date: either before or after 2010
2. Enrollment type: whether the program is mandatory or voluntary
3. Geographic reach: whether the program is state-wide, or in limited regions

We discuss how these program features affect the ultimate evaluation design in detail below, and summarize these decisions in Table V.1.

Table V.1. Design options by program feature

Program start date	Enrollment type	Geographic reach	Preferred evaluation design
2010 or later	Mandatory	Limited	DID with in- or out-of-state comparison group
		State-wide	DID with out-of-state comparison group
	Voluntary	Limited	DID* with in- or out-of-state comparison group
		State-wide	DID* with out-of-state comparison group, or RCS in-state
Before 2010	Mandatory	Limited	RCS with in- or out-of-state comparison group
		State-wide	RCS with out-of-state comparison group
	Voluntary	Limited	RCS with in- or out-of-state comparison group
		State-wide	RCS with in- or out-of-state comparison group

DID = difference-in-differences; DID* = difference-in-differences with Bloom correction; RCS = repeated crosssectional.

Program start date. Whether the program started in 2010 or later will affect whether data prior to MLTSS program implementation are available, and in turn, whether we will be able to implement a DID or RCS design. For MLTSS programs that started prior to 2010 (16 of the 34 programs listed in Table II.1), we will be unable to use claims data prior to the program start date because these data do not have the relevant HCBS taxonomy codes needed to construct comparison groups.[14] Therefore, for these 16 programs, we will only have data after MLTSS implementation, and we will need to rely on a RCS design to evaluate these programs. A DID design may be possible for the 18 programs that started in 2010 or later.

Enrollment type and geographic reach. Both enrollment type (mandatory or voluntary) and geographic reach affect whether or not the program covers all LTSS users in a particular state, which in turn affects both the over-arching design (DID vs. RCS) as well as the location of the comparison group. As mentioned above, an in-state comparison group will not be possible when the program is mandatory state-wide, because these programs would not have a within state pool of eligible FFS LTSS recipients to use as a comparison group. However, even in cases

[14] MAX began applying the HCBS taxonomy to FFS HCBS claims submitted under 1915(c) waivers beginning in 2010. For the final evaluation, we will replicate the methodology that MAX uses to apply the HCBS taxonomy to encounter claims in all years, and to FFS waiver claims in 2009 for programs that started in 2010 in order to have pre-period data for the 2010 programs. Specifically, only Tennessee had an MLTSS program that began in 2010, so this is the only program for which we will need to replicate the HCBS taxonomy in 2009 for pre-period data. This will align with the approach used in the interim evaluation (Libersky et al. 2018).

where the program is voluntary and/or only available in limited regions, an in-state comparison group may not be the best choice. Factors that will affect our decision about whether to use an in-state versus an out-of-state comparison group include the sample sizes for MLTSS enrollees and FFS LTSS users, and the geographic and demographic similarity of the two groups.[15] As general guidelines, in order to use an in-state comparison group we will require at least 30 percent of eligible LTSS beneficiaries in the state to be enrolled in FFS (overall and within the key subgroups), as well as a reasonable amount of overlap in the geographic (urban/rural) and demographic features of the two groups. We will assess whether an in-state or out-of-state comparison group is a better choice carefully, on a program-by-program basis, once we have further information on the enrollment characteristics for each program.

Voluntary MLTSS programs pose an additional challenge in DID designs, because all eligible beneficiaries in a particular region that is covered by the program do not actually participate. If we define our MLTSS group to be only those who opt in to the program, we introduce a potential source of selection bias if factors related to both program participation and outcomes of interest are unobserved. In order to protect against this source of bias, we can avoid the selection issue by defining our MLTSS group to be all LTSS users who live in regions covered by MLTSS programs, regardless of whether or not they opt in to the program. This design will produce estimates of the impact of *access* to MLTSS programs on outcomes. We can then apply corrections to these estimates, in order to obtain estimates of the impact of program *participation*, which is the impact of interest. One example of such a correction is known as the Bloom correction (Bloom 1984). This approach will be most valid in cases where the proportion of eligible beneficiaries who do not opt in to the program is relatively small. If opting out of the MLTSS program is common, we may instead choose to conduct an in-state RCS evaluation.

The approach outlined above (DID with Bloom correction) can be used for voluntary programs starting after 2010, regardless of the geographic reach. Note that if these programs are statewide, the comparison group must be out-of-state, even though there are FFS beneficiaries within the state. This is because there is no geographic delineation that separates MLTSS and FFS participants within the state. For these programs, we will use data on MLTSS penetration rates and beneficiary characteristics in order to make a decision as to whether an in-state RCS or an out-of-state DID analysis (with Bloom correction) represents a stronger evaluation design.

Voluntary programs for which pre-period information is not available (those that began before 2010) are particularly susceptible to selection bias, because we have no way of employing a DID design to remove differences between the MLTSS and FFS groups that existed prior to program participation. This will be a limitation of our evaluation for these programs.

2. Study sample

The quality of our evaluation is contingent on the sample selection process. The primary goal of the sample selection process is to accurately approximate the counterfactual: the outcomes that would have been observed on MLTSS users, had they never enrolled in the

[15] For example, if a program is not statewide, but it includes the most populous counties in the state and there are only a small number of in-state FFS LTSS users who are not enrolled in MLTSS, those LTSS users would likely not be a strong comparison group. Similarly, if the program is statewide but voluntary, but the vast majority of LTSS eligible beneficiaries opt into MLTSS, we may be able to find a stronger out-of-state comparison group.

program. Because participation in MLTSS is nonrandom, we need to construct a comparison group that appears similar to the MLTSS enrollees on key observable characteristics that affect MLTSS enrollment and outcomes. The remainder of this section details the steps required for sample selection, for both DID and RCS designs.

a. Choosing comparison states

To create out-of-state comparison groups, we will select states with FFS LTSS that are as similar as possible to the MLTSS states on a number of relevant factors. Similar to the approach used for the interim evaluation in Tennessee (Libersky et al. 2018), we will select FFS states to include for the comparison group based on data availability, geographic proximity, demographic similarity, and comparable values for environmental measures of LTSS supply, demand, and Medicaid LTSS rebalancing indicators, including:

1. HCBS spending as a share of total LTSS spending for (1) all populations, (2) adults over age 65 and people with physical disabilities, and (3) people with developmental disabilities (LTSS Expenditure Reports)
2. Percent of adults age 21 or older with an ADL-limiting disability and income at or below 250 percent of the federal poverty level receiving Medicaid or other government assistance health insurance (based on AARP LTSS Scorecard, constructed from American Community Survey)
3. Number of home health/personal care aides per 100 people age 18 and older with an ADL-limiting disability (based on AARP LTSS Scorecard, constructed from American Community Survey)
4. Number of nursing facility beds per 100 people age 18 and older with an ADL-limiting disability (constructed from American Community Survey and Area Health Resource File)
5. Percentage of dually eligible beneficiaries enrolled in Dual Special Needs Plans (Medicare Advantage/Part D Enrollment and Contract Data: Special Needs Plan Data)

If there are multiple states available that meet our comparison group selection criteria, we will consider constructing a comparison group from multiple states. Doing so could provide more robustness to our estimates, as our impact estimates would be less reliant on similarity between the MLTSS enrollees and FFS users in a single state.

b. Study eligibility

In order to select the study sample used to evaluate each program, we first need to define a set of eligibility criteria in a way that can be applied consistently to the MLTSS and comparison groups. In order to do so, we will attempt to replicate MLTSS enrollment requirements for each program as closely as possible in the claims data. The requirements and definition of study eligibility will differ between states and programs.

When possible, we will use the type of plan and plan ID recorded in the MAX/AlphaMAX/TAF eligibility files to identify MLTSS enrollees. We will confirm with each state that the plan IDs that we are using were active during the study period. However, the plan ID/type recorded in the Medicaid claims file may not contain all the information needed to identify enrollees in certain MLTSS programs (for example, if the same plan covers both

MLTSS and managed care for medical services alone for people who do not qualify for LTSS). In these cases, we will contact the state to request an individual-level finder file that will allow us to link with MAX/Alpha-MAX/TAF for the identification.

Because we do not have information on level of care and functional status from claims data, it would be difficult for us to accurately identify comparison groups for MLTSS enrollees with a less than institutional level of care and/or no LTSS need. Without this information, there would be serious concerns that there could be unobserved differences between MLTSS enrollees and comparison beneficiaries that would impact our estimates for these programs. Due to these issues, we will limit our evaluation to individuals who meet an institutional level of care, thereby excluding individuals who meet a less than institutional level of care or gain eligibility through dual status and may not use LTSS during our study period. For MLTSS programs that require an institutional level of care, we can include all MLTSS enrollees as part of the study population. For MLTSS programs that allow less than institutional level of care or no LTSS need, we will use several sources to identify beneficiaries meeting an institutional level of care. Because MAX/AlphaMAX data do not include information on level of care, and the quality of the level of care fields available in TAF is unknown, we will assess institutional level of care based on (1) program characteristics described in waiver documentation, (2) TAF data, where available, and/or (3) finder files obtained directly from states.

Our comparison groups will consist of FFS 1915(c) waiver enrollees, who by definition meet an institutional level of care, and individuals residing in institutions. Because some MLTSS programs cover a different range of services, we will tailor the eligibility criteria to the list of services covered in each MLTSS program. For example, if certain types of institutions are excluded from capitation in the MLTSS program, we will not include individuals residing in those types of institutions as part of our comparison group. Other factors that will define comparison group eligibility include age and type of disability covered in the MLTSS program, Medicare status, level of LTSS need, and services excluded from capitation.

c. Beneficiary selection

Propensity score matching. We will use propensity score matching to determine the final matched comparison group (Rosenbaum and Rubin 1983). This approach allows for an approximation of an experimental design by assuming that the decision to participate is random, conditional on a set of observable characteristics.

The propensity score is estimated from a regression model fit to the sample of beneficiaries that includes both MLTSS enrollees and the potential comparison group who meet the inclusion/exclusion criteria. The dependent variable in the propensity score model is MLTSS enrollment, and the independent variables include factors hypothesized to be related to participation in the MLTSS program and outcomes. We will fit a separate propensity score model for each MLTSS program; this allows the propensity score model to be tailored to each program, which should result in better program-level balance between the characteristics of the MLTSS and matched comparison groups.

Propensity score matching is a widely used and effective method of identifying a matched comparison group based on a set of observed variables. In designing our matching procedure for each program, we will group observed variables into three levels based on how important we believe they are to ensure a good balance between the MLTSS and comparison groups. For the highest-priority variables (such as dual status, and other key subgroups), we will use exact matching: only allowing an MLTSS enrollee to be matched to a potential comparison beneficiary with the same value for this variable. For the next level of variables, we will apply a caliper: only allowing an MLTSS enrollee to be matched to a potential comparison beneficiary if their values of the variable are within a certain prespecified range. Applying calipers is a matching method commonly used to reduce the difference in covariate distribution between two populations (Stuart 2010). The remaining variables will be balanced through their contribution to the propensity score.

Having different start dates within a given program causes a challenge: since beneficiaries in the comparison group do not have a specific enrollment date (by definition, they are never enrolled), it is not immediately clear how to define their matching variables that vary over time. In these cases we will employ a common technique that we call “matching with replicates,” in which a different set of matching covariates is constructed for every possible month of enrollment, for each comparison subject. There are various ways to implement this method (Chew et al. 2017), however Mathematica staff have developed a strategy that is superior to existing methods, both in terms of the covariate balance between the treatment and matched comparison groups, and the efficiency of the resulting impact estimates.[16] The method is based on optimal matching (Rosenbaum 1989; Hansen and Klopfer 2006), but implements constraints that are appropriate for the structure of the problem. It has been used successfully on other Mathematica evaluations for the Centers for Medicare and Medicaid Services, including the Health Care Innovation Awards (HCIA) Round Two and Transforming Clinical Practice Initiative (TCPI).

Beneficiary selection in DID designs. For DID designs, we will construct a panel of eligible MLTSS enrollees and matched comparison FFS LTSS users. We will construct this panel in 12-month increments relative to the start of the program. This will include one 12-month period prior to the start of the program, and as many 12-month periods as possible after the start of the program, up until the end of the study period. For each year, we select our MLTSS group as all individuals who meet our study eligibility criteria (V.A.2.a) that live in areas covered by MLTSS programs. For voluntary programs, this will include individuals who do not choose to enroll in the program; as discussed above, we will account for this feature by applying a Bloom correction to our impact estimates.

The comparison group for each 12-month study period will be chosen from FFS LTSS users in regions not covered by an MLTSS program who meet study eligibility criteria. From this pool of beneficiaries, we will select a matched comparison group using propensity score matching. Since the propensity score model will be applied to both pre-intervention and post-intervention periods, we will only include exogenous covariates (those that we do not believe can be affected by program participation) in our propensity score model. Examples of such covariates include

[16] The paper detailing this method is currently in preparation, but the approach has been presented at various statistical conferences over the past year.

demographic information and chronic conditions, as measured by the Chronic Illness and Disability Payment System (CDPS) diagnostic classification system.

Beneficiary selection in RCS designs. An RCS design will be used in cases where we do not observe data from before the start of the program, because the program started before our study period begins (2010). The beneficiary selection process differs from the DID in that, rather than constructing a panel whose members could change from year to year, beneficiaries will be assigned to either the MLTSS or comparison group once, and they will remain in that group for the duration of the study, following an intent-to-treat approach. Thus, each MLTSS enrollee will only be matched once, as opposed to being matched for each study period in the DID design. We will restrict our MLTSS population to enrollees who meet the study eligibility criteria, as discussed above. Similarly, we will only allow comparison beneficiaries to be matched if they meet the study eligibility criteria at the time of enrollment for the person to whom they are matched. As in the DID designs, we will only include exogenous covariates in our matching model, such as demographics and chronic conditions.

Note that there are two subgroups of enrollees in the RCS design: those who were already enrolled in the program at the start of the study period (January 1, 2010), and those who enroll later. The latter subgroup introduces a complication in the case of a voluntary program with no geographic distinction between regions covered and not covered by MLTSS programs (that is, an in-state comparison group for a state-wide program). The complication arises because an eligible beneficiary who is not enrolled at the start of our study (and thus could be a potential comparison member) could later enroll in the program (and thus serve as a member of the MLTSS group). In order to avoid this complication without biasing our estimates, we will exclude all beneficiaries who were not enrolled at the start of our study period from the MLTSS group for programs where we will have in-state comparison group for a state-wide program. Thus, the study sample for these programs will be defined at a single point in time, the start of our study period. We will follow an intent-to-treat approach, so the members of our sample will remain in the group to which they were originally assigned, regardless of whether or not they drop out or enroll late in the program.

For other programs (mandatory programs, or voluntary programs that have geographic distinction between regions that are and are not covered by MLTSS programs), we will include beneficiaries who enroll after the start of the study period provided they meet the study eligibility criteria.

Assessing the quality of matching. Using matching to select a comparison group will produce unbiased estimates if two assumptions are met: (1) the set of observable characteristics used in the matching procedure includes all factors related to both MLTSS enrollment and the outcomes; and (2) enrollees and comparison group members are "balanced" on observable characteristics conditional on their propensity score within each stratum—that is, for each enrollee, there must be a matched comparison group member(s) similar to the participant on observed characteristics (Rosenbaum and Rubin 1983). The first assumption is untestable, but is assumed to be true in this evaluation design. To determine whether the latter condition is met, we will perform several statistical tests. Following Stuart (2010), we will examine differences in means and standardized differences of the variables used in the matching process. We will assess whether standardized bias for each variable is less than 0.25, a commonly used threshold to ensure balance, as well as a stricter threshold of 0.10. This assessment will indicate whether our matching procedure produces a comparison group that is similar to the MLTSS group for each variable in our model. If standardized differences above 0.25 still exist after matching, we will consider including calipers on that variable in order to improve its balance or, if needed, an alternative matching approach. We will present standardized differences for all matching variables for each evaluated program, using graphical displays to efficiently convey large amounts of information.

3. Program-level regression analysis

We will estimate program impacts using regression techniques. This section describes the structure of the data and models, as well as how we plan to handle certain data complexities.

a. Data structure

The analytic dataset for our regression models will be constructed at the beneficiary-month level. This means that we will measure both outcomes and predictors (demographics and CDPS) for every month that the person remains in the sample. CDPS will be constructed on a rolling monthly basis, using the prior 12-months relative to the analysis month, to ensure relatively stable measures over time. Using monthly data, as opposed to annual, solves complications that arise when individuals are only observed for part of a year. This will occur frequently, whenever beneficiaries newly enroll, die, move out of state, or lose Medicaid eligibility midway through a study year. Outcomes observed over partial years are inherently more noisy than full-year observations, and may represent under-counts of key outcomes compared to what they would have been had the beneficiary been observed for the full year. Constructing the file at the monthly level mitigates these concerns.[17]

b. Model type

We will analyze 11 outcome measures related to service use and quality of care, each defined at the monthly level. For the 10 outcome measures that are dichotomous (binary) variables (any HCBS use, any ILTC use, any round-the-clock services use, any day services use, any home-delivered meals use, any home-based services use, any caregiver support services use, any equipment, technology, or modifications use, minimizing institutional length of stay, and any potentially avoidable hospitalization), we will use logistic regression models. For the one outcome measure that is a count variable (number of hospital days), we will consider several modeling approaches, including linear regression, negative binomial regression, zeroinflated models, and two-part models. We will make the final determination of the approach after analyzing the distribution of the variable and the residuals from fitting each model.

c. Model specifications

The specification of the regression model will differ between the DID and RCS designs.

[17] We do assume that each beneficiary is observed for a full month, whenever they are observed. However, this assumption is less strict than assuming each year is fully observed, because we will have a much greater proportion of time periods that are fully observed when we organize data at the monthly level.

Model specification for DID designs. For DID designs, we will have panel data for two groups (MLTSS and comparison), including 12 months of pre-intervention data and several years' worth of monthly post-intervention data. The regression models will take the following form:

$$g(\mu_{it}) = \alpha + \beta X_{it} + \gamma MLTSS_i + \delta_t + \theta_t MLTSS_i * post_t$$

In this model, *i* indexes the beneficiary and *t* indexes the study month, μ_{it} is the expected value of the outcome for beneficiary *i* during month *t*, and $g()$ is the generalized linear model link function (for example, logit for logistic regression). Also, X_{it} is a vector of covariates (for example, demographic characteristics and CDPS) for beneficiary *i* during month *t*, $MLTSS_i$ is the indicator that beneficiary *i* is in the area that implemented the MLTSS program, and $post_t$ is an indicator that month *t* is in the post-period. The Greek letters are all parameters to be estimated.

The primary parameters of interest in this model are the coefficients of the interaction terms, θ_t. These parameters can be interpreted as the expected change in the outcome (on the scale of the link function) comparing the MLTSS to comparison groups during program month *t*, subtracting out differences that existed at baseline and holding all other covariates constant. These parameters may be difficult to interpret, for two reasons. First, they are on the scale of the link function (for example, log odds ratios for logistic regression), which is not a natural way to think about impact estimates. Second, there will be many of these interaction terms (one for each program month), each of which will be relatively noisy (because it corresponds to a small time period). In order to present more interpretable results, our final impact estimates will be calculated as annual marginal effects on the natural scale of the outcome (the probability scale for binary outcomes). These marginal effects will be calculated by marginalizing (averaging) over the distribution of the MLTSS group, as well as overall months in a given study year. The resulting impact estimates will be annual estimates of the average treatment effect on the treated (ATT). For example, we will present estimates of the impact of MLTSS participation as a percentage point change in average monthly ILTC use (a binary outcome). These impacts can be interpreted as the expected difference in the proportion of beneficiaries with any ILTC use during any month in a given program year, comparing MLTSS enrollees in the program to their expected outcomes had they never received the program. We will present these impact estimates for each program year during the study period.

Other parameters in the model can be interpreted as follows. The intercept α is the expected outcome (on the scale of the link function) for the comparison group in the pre-intervention period when all covariates are equal to 0; the vector of coefficients β represents the effect of covariates on the outcome, γ is the difference between the MLTSS and FFS groups in the preintervention period; and δ_t accounts for a secular time trend assumed to be identical for both groups. We will account for within-beneficiary clustering of outcomes using a clustered sandwich estimator of the variance (Wooldridge 2002).

The key assumption of DID models is known as the "parallel trends" assumption: that in the absence of the program, any patterns in outcomes (that is, changes in outcomes over time) that we observe in the comparison group would have been observed in the MLTSS group, had those beneficiaries not enrolled in the program. This assumption is more likely to hold when we use an in-state comparison group, when the state-specific environment is constant between the two groups. Though the parallel trends assumption is inherently untestable, we will examine patterns of monthly outcomes during the pre-enrollment period, in order to assess the credibility of this assumption. Observing parallel trends during the pre-enrollment period does not prove that these trends will continue during follow-up, but it does provide some supporting evidence in favor of the assumption.

Model specification for RCS designs. In RCS designs we will have longitudinal data of program participants and their matched comparisons over the study period. The key distinction from the DID design is that outcomes are only measured in the post-intervention period. The regression models will take the following form:

$$g(\mu_{it}) = a_i + \beta X_{it} + \delta_t + \theta_t MLTSS_i$$

As in the DID model, *i* indexes the beneficiary and *t* indexes the program month, except that here time is represented in calendar time relative to the start of our study period, as opposed to time since the start of the program (because program start typically does not fall within our study period).[18] μ_{it} is the expected value of the outcome for beneficiary *i* during month *t*, $()$ is the generalized linear model link function, X_{it} is a vector of covariates (for example beneficiary demographics and CDPS), and $MLTSS_i$ is the indicator that beneficiary *i* is a participant in an MLTSS program. The model differs from the DID model above in that we do not have data from a pre-enrollment period, so there is no post-period indicator or interaction term.

The primary parameters of interest in this model are the coefficients of the $MLTSS_i$ indicator, θ_t, which represent the expected difference in outcomes between the MLTSS and FFS groups (on the scale of the link function), during month *t*, holding all covariates constant. As above, we will present annual ATT impacts by calculating marginal effects on the natural scale of the outcome (for example, percentage point changes for binary outcome measures), calculated over the distribution of the MLTSS group and all months in the year. These impacts can be interpreted exactly as the impacts from the DID model are interpreted. The only difference is that, because we are unable to subtract out any differences in outcomes that existed prior to the program, the impacts are potentially more susceptible to selection bias if the MLTSS and FFS groups differ on unobserved confounders.

Other parameters in the model include the covariate effects β and the monthly intercepts δ_t, which account for a secular time trend in the comparison group. We also include random effects α_i to account for within-beneficiary clustering (this was less practical in the DID model due to the panel nature of the data). Note that we have intentionally omitted the overall intercept (α in the DID model), which allows δ_t to be interpreted as the expected outcome in the comparison group during month *t*, when $X_{it} = 0$.

[18] We may have programs that started after 2010 that we decide to evaluate using an RCS design. In this case, we have the option of defining time relative to the start of the program. We also would be able to include many more covariates as part of our covariate vector X_{it}, including baseline measures of key outcome variables.

4. Meta-analysis

a. Goals and framework

We will synthesize findings across programs, outcomes, and time points using meta-analysis techniques. The meta-analysis will serve two purposes. First, it will allow us to summarize findings across different programs and time points, providing, for example, the average impact of MLTSS on each outcome, among all the programs we evaluate. Second, the meta-analysis will allow us to understand how different program features or state-level LTSS environment factors relate to outcomes. Program features that we will consider for this analysis include those described in Section II, such as mandatory or voluntary enrollment, and whether the program is state-wide or in limited regions. Environmental factors may include those considered when choosing out-of-state comparison groups, for example, the number of home health/personal care aides per 1,000 people over age 65.

In contrast to the regression models discussed in Section V.A.3, the outcomes in these models will not be beneficiary-specific. Rather, the dependent variable used in the meta-analysis will be the impact estimates for each program, outcome, and time point that result from the program-specific impact analyses. The model will be fit in a Bayesian hierarchical framework, which is uniquely well-suited for synthesizing results across different dimensions of the model (outcome types, programs/states, and time). The hierarchical model decomposes the residual uncertainty into components related to each of these three dimensions. To the extent that it is supported by the data, outcomes that correspond to the same program, outcome, or time point will inform one another, a phenomenon known as “borrowing strength” or “shrinkage.” This approach increases statistical precision, providing more accurate estimates of model components. The Bayesian approach also provides an easy, natural method for incorporating the uncertainty of each impact estimate measure.

b. Model specification

The meta-analysis analysis takes the program-specific impacts of MLTSS on service use and quality of care as inputs, and relates them to program features. The regression model takes the following form:

$$\theta_{jkt} = \alpha + \beta_j X_k + \gamma_j Z_k + \varepsilon_{jkt}$$

$$\varepsilon_{jkt} = b_j + c_k + d_t^{DID} DID_k + d_t^{RCS} RCS_k + e_{jkt}$$

In this model, we let *j* index outcomes, *k* index MLTSS programs, and *t* index time. The variable θ_{jkt} is the estimated impact of MLTSS program *k* on outcome *j* at time *t*, X_k are program features that correspond to program *k*, and Z_k are environmental features that correspond to the state in which program *k* is implemented. ε_{jkt} is the residual that remains after taking into account the program features, which we decompose into an outcome-specific component (b_j), a program-specific component (c_k), and a time-specific component ($d_t^{DID} DID_k + d_t^{RCS} RCS_k$, where DID_k and RCS_k are the indicators that program *k* was evaluated using a DID or RCS design, respectively). Note that the model controls for time differently for programs evaluated using DID or RCS, as time is defined relative to program start in DID designs, but relative to our study period for RCS designs.

The parameter $\alpha + b_j$ represents the expected impact of MLTSS on outcome *j* for a program with all program features (X_k) coded as zero. Other parameters of interest are β_j and γ_j, which reflect the association of each program or environmental feature on outcome *j*. More specifically, we can interpret the elements of the vector β_j as the expected change in the impact estimate on outcome *j* during any particular year for every 1-unit change in the corresponding program feature (holding all else constant), and the elements of γ_j can be interpreted as the expected change in the impact estimate on outcome *j* for every one-unit change in the corresponding environmental feature.

An important aspect of this approach is that we include the impact estimates from all outcomes in a single framework. These impact estimates not only refer to different outcomes, but they could fall on very different scales, as we have both binary and continuous outcomes. In order for this approach to be sensible, we will first standardize all impact estimates so they are on a consistent scale. We will also consider a hierarchical prior specification for the outcome specific intercept, to allow for systematic differences in impacts between outcomes of various types or substantive categories (for example, utilization versus quality of care). If we do not feel it appropriate to combine certain outcomes in a single model, we may decide to separate outcomes into two models. These decisions, along with specifics on prior specification, will be made once we know the exact number of impact estimates that will enter the model, which will flow in turn from the number of states with available data that pass our assessments of data quality.

B. Methods for examining access, beneficiary experience, and quality of life (research questions 3 and 4)

As described in Section IV, we will compare NCI-AD indicators for MLTSS and FFS population. If possible, we will use beneficiary-level NCI-AD data for all available states; otherwise, we will use state-specific summary reports from NASUAD. Where possible, the final report will group findings by MLTSS and FFS models, and identify major differences in surveyed populations and covered benefits (as in comprehensive versus limited benefit programs) that affect results across states.

In addition to this descriptive analysis, we will compare survey responses across states using a Bayesian hierarchical regression model. This model will include mean responses to all survey items in the same model, much like the meta-analysis described in Section V.A.4 includes all outcomes in the same model. The Bayesian framework offers the same advantages as it did in that model: it allows us to specify hierarchical priors so that the model can borrow strength across survey items and states, and allows for a convenient framework to incorporate uncertainty around survey means. The regression model will include both MLTSS and FFS states, allowing us to estimate the impact of a state having an MLTSS program on outcomes.

The regression model will take the following form:

$$Y_{js} = \alpha + \psi_j MLTSS_s + \beta_j X_s * MLTSS_s + \gamma_j Z_s + \varepsilon_{js}$$

$$\varepsilon_{js} = b_j + c_s + e_{js}$$

In this model, Y_{js} is the mean response for survey item *j* in state *s*, $MLTSS_s$ is the indicator that state *s* is an MLTSS state, X_s are program-specific features that apply to states with MLTSS programs, and Z_s are state-specific environmental features. The parameter ψ_j represents the overall impact of MLTSS on survey item *j*, whereas β_j and γ_j reflect the associations between program-specific features and state-specific environmental features, respectively, with outcome *j*. As in the model for service use and quality of care, we decompose the residual into an item specific component (b_j) and a state-specific component (c_s). We will apply hierarchical priors to all regression parameters, as appropriate.

C. Methods for examining MLTSS per-user expenditures (research question 5)

As in Section V.B, we will examine state-level MLTSS spending both through descriptive analysis and regression modeling. First, we will present state-level descriptive trends in (1) MLTSS spending by service category, and, if appropriate, (2) MLTSS spending per user or enrollee for all MLTSS states from 2013–2017. As described in Section IV, we will use aggregate annual summary data collected for CMS's LTSS expenditure reports and Medicaid's managed care enrollment report to produce these descriptive trends for states with MLTSS. Combining information from these two different sources will give us a rough estimate of per enrollee expenditures.

We will conduct a state-level regression analysis in order to gain a better understanding of the variation in MLTSS spending across states. The model will be fit using a Bayesian hierarchical regression model, similar to the models described in Sections V.A.4 and V.B. The regression model will take the following form:

$$Y_{st} = \alpha + \beta X_s + \gamma Z_s + \varepsilon_{st}$$

$$\varepsilon_{st} = c_s + d_t + e_{st}$$

In this model, Y_{st} are the average expenditures per MLTSS user in state *s* at time *t*, X_s are program features for state *s* (aggregated to the state level), and Z_s are state-specific environmental features. Once again, we decompose the residual into a state-specific component and a time-specific component, for which we will specify appropriate hierarchical priors. The parameter α represents the average per-user expenditures for a state with all program features set to zero, and the parameters β and γ are the effects of each program or environmental feature on the expenditures for that state.

VI. LIMITATIONS

One of the major limitations of our evaluation is that we will only be to conduct a rigorous evaluation for the claims-based outcome measures in the states with sufficient data availability and quality. Because of the historically poor quality of the encounter data in MSIS and data quality issues with the transition to T-MSIS, our evaluation will likely be limited to a few select states that meet our data quality thresholds for inclusion. This will limit our ability to generalize findings to other MLTSS states. It also may place constraints on the meta-analysis. If we can only evaluate one or two programs, a cross-program meta-analysis would likely result in very uncertain estimates (that is, high estimated standard errors), which would limit what we are able to take away from such an analysis.

For the states that we are able to include in the evaluation of service use and quality of care with a DID design, our estimates are dependent on the credibility of the parallel trends assumption, as well as the assumption that the treatment and comparison groups are similar on both observed and unobserved characteristics. If these assumptions do not hold, our results will be biased. For the states that we are able to include in the evaluation with a RCS design, we are unable to account for preexisting differences between the treatment and comparison groups. In addition, we will be limited in the characteristics that we are able to use for matching.

While beneficiary access and experience are key outcomes for understanding how MLTSS models compare to FFS LTSS, NCI-AD data are limited to a few states, and at most, two time periods. In addition, the two time periods are only one year apart, limiting our ability to observe trends. For our descriptive analysis of MLTSS spending, we will be limited to analyzing within-state trends. Further, because available data varies across states, we will not be able to report all outcomes for all MLTSS states.

Despite these limitations, this evaluation will expand and improve on findings from the interim evaluation and will identify areas for future analysis. Importantly, we will use all available data to summarize findings across states and by selected program features by conducting a meta-analysis.

This page has been left blank for double-sided copying.

REFERENCES

Bloom, Howard S. "Accounting for No-Shows in Experimental Evaluation Designs." *Evaluation Review*, vol. 8, no. 2, 1984, pp. 225–46.

Centers for Medicare & Medicaid Services Medicare-Medicaid Coordination Office (MMCO). "Data Analysis Brief: Managed Care Enrollment Trends among Medicare-Medicaid Beneficiaries and Medicare-only Beneficiaries, 2006 through 2016." November 2017. Available at https://www.cms.gov/Medicare-Medicaid-Coordination/Medicare-and-Medicaid-Coordination/Medicare-Medicaid-Coordination-Office/Downloads/ManagedCareEnrollmentTrends2006-2016Data.pdf. Accessed September 12, 2018.

Chew, Rob, Kasey Jones, Mahin Manley, Allison Witman, Chris Beadles and Ann Larson. "Rollmatch: Rolling Entry Matching. R package version 1.0.0." 2017.

Hansen, Ben B., and Stephanie Olsen Klopfer. "Optimal Full Matching and Related Designs Via Network Flows." *Journal of computational and Graphical Statistics*, vol. 15, no. 3, 2006, pp. 609–627.

Irvin, Carol, Jessica Heeringa, Debra Lipson, Jenna Libersky, Carey Appold, Maggie Colby, Vivian L.H. Byrd, Maggie Colby, Julia Baller, and Katherine Bradley. "Medicaid 1115 Demonstration Evaluation Design Plan. Washington, DC: Mathematica Policy Research, May 15, 2015.

Libersky, Jenna. "What We Know—and Don't Know—About Access to Long-Term Services and Supports Programs." Blog post. Washington, DC: Mathematica Policy Research, May 30, 2018. Available at https://www.mathematica-mpr.com/commentary/what-we-know-anddont-know-about-access-to-long-term-services-and-supports-programs. Accessed August 24, 2018.

Libersky, Jenna, Su Liu, Laura Ruttner, Anna Collins, Jonathan Geller, and Carol Irvin. "Managed Long-Term Services and Supports Design Supplement: Interim Outcomes Evaluation." Baltimore, MD: Centers for Medicare & Medicaid Services, June 2017.

Libersky, Jenna, Su Liu, Laura Turoff, Jonathon Gellar, Debra Lipson, Anna Collins, Jiaqi Li, and Carol Irvin. "Medicaid 1115 Demonstration Interim Evaluation Report: Managed Long Term Services and Supports." Washington, DC: Mathematica Policy Research, January 31, 2018. Available at https://www.medicaid.gov/medicaid/section-1115demo/downloads/evaluation-reports/mltss-interim-eval-report.pdf. Accessed September 21, 2018.

NASUAD and Human Services Research Institute (HSRI). "NCI-AD Consumer Survey: 2015-2016 National Results." Washington, DC: NASUAD and Cambridge, MA: HSRI, May 2017. Available at https://nci-ad.org/upload/reports/NCI-AD_20152016_National_Report_FINAL.pdf. Accessed August 23, 2018.

NASUAD and HSRI. "NCI-AD Consumer Survey: 2016-2017 Colorado Results." Washington, DC: NASUAD and Cambridge, MA: HSRI, March 2018a. Available at https://nciad.org/upload/state-reports/NCI-AD_2016-2017_CO_state_report_FINAL.pdf. Accessed September 5, 2018.

NASUAD and HSRI. "NCI-AD Consumer Survey: 2016-2017 Indiana Results." Washington, DC: NASUAD and Cambridge, MA: HSRI, May 2018b. Available at https://nciad.org/upload/state-reports/NCI-AD_2016-2017_IN_state_report_FINAL.pdf. Accessed September 5, 2018.

NASUAD and HSRI. "NCI-AD Consumer Survey: 2016-2017 Kansas Results." Washington, DC: NASUAD and Cambridge, MA: HSRI, May 2018c. Available at https://nciad.org/upload/state-reports/NCI-AD_2016-2017_KS_state_report_FINAL.pdf. Accessed September 5, 2018.

NASUAD and HSRI. "NCI-AD Consumer Survey: 2016-2017 Maine Results." Washington, DC: NASUAD and Cambridge, MA: HSRI, May 2018d. Available at https://nciad.org/upload/state-reports/NCI-AD_2016-2017_ME_state_report_FINAL.pdf. Accessed September 5, 2018.

NASUAD and HSRI. "NCI-AD Consumer Survey: 2016-2017 Minnesota Results." Washington, DC: NASUAD and Cambridge, MA: HSRI, April 2018e. Available at https://nciad.org/upload/state-reports/NCI-AD_2016-2017_MN_state_report_FINAL_508.pdf. Accessed September 5, 2018.

NASUAD and HSRI. "NCI-AD Consumer Survey: 2016-2017 Mississippi Results." Washington, DC: NASUAD and Cambridge, MA: HSRI, February 2018f. Available at https://nci-ad.org/upload/state-reports/NCI-AD_2016-2017_MS_state_report_FINAL.pdf. Accessed September 5, 2018.

NASUAD and HSRI. "NCI-AD Consumer Survey: 2016-2017 Nevada Results." Washington, DC: NASUAD and Cambridge, MA: HSRI, May 2018g. Available at https://nciad.org/upload/state-reports/NCI-AD_2016-2017_NV_state_report_FINAL.pdf. Accessed September 5, 2018.

NASUAD and HSRI. "NCI-AD Consumer Survey: 2016-2017 New Jersey Results." Washington, DC: NASUAD and Cambridge, MA: HSRI, April 2018h. Available at https://nci-ad.org/upload/state-reports/NCI-AD_2016-2017_NJ_state_report_FINAL_2.pdf. Accessed September 5, 2018.

NASUAD and HSRI. "NCI-AD Consumer Survey: 2016-2017 Ohio Results." Washington, DC: NASUAD and Cambridge, MA: HSRI, April 2018i. Available at https://nciad.org/upload/state-reports/NCI-AD_2016-2017_OH_state_report_FINAL_2.pdf. Accessed September 5, 2018.

NASUAD and HSRI. “NCI-AD Consumer Survey: 2016-2017 Oregon Results.” Washington, DC: NASUAD and Cambridge, MA: HSRI, January 2018j. Available at https://nciad.org/upload/state-reports/NCI-AD_2016-2017_OR_state_report_FINAL.pdf. Accessed September 5, 2018.

NASUAD and HSRI. “NCI-AD Consumer Survey: 2016-2017 Pennsylvania Results.” Washington, DC: NASUAD and Cambridge, MA: HSRI, March 2018k. Available at https://nci-ad.org/upload/state-reports/NCI-AD_2016-2017_PA_state_report_FINAL.pdf. Accessed September 5, 2018.

NASUAD and HSRI. “NCI-AD Consumer Survey: 2016-2017 Tennessee Results.” Washington, DC: NASUAD and Cambridge, MA: HSRI, April 2018l. Available at https://nciad.org/upload/state-reports/NCI-AD_2016-2017_TN_state_report_FINAL.pdf. Accessed September 5, 2018.

NASUAD. “Participating NCI-AD States.” Washington, DC: NASUAD, updated 2018. Available at: https://nci-ad.org/states/. Accessed August 23, 2018.

Rosenbaum, Paul R. “Optimal Matching for Observational Studies.” *Journal of the American Statistical Association*, vol. 84, no. 408, 1989, pp. 1024–1032.

Rosenbaum, P.R., and D.B. Rubin. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika*, vol. 70, no, 1, April 1983, pp.41–55.

Saucier, P., J. Kasten, B. Burwell, and L. Gold. “The Growth of Managed Long-Term Services and Supports Programs: A 2012 Update.” Baltimore, MD: Truven Health Analytics, July 2012. Available at https://www.medicaid.gov/medicaid-chip-program-information/by-topics/delivery-systems/downloads/mltssp_white_paper_combined.pdf. Accessed September 21, 2018.

Verdier, James, Alexandra Kruse, Rebecca Sweetland Lester, Ann Mary Philip, and Danielle Chelminsky. “State Contracting with Medicare Advantage Dual Eligible Special Needs Plans: Issues and Options.” Integrated Care Resource Center Technical Assistance Tool, November 2016. Available at http://www.integratedcareresourcecenter.com/PDFs/ICRC_DSNP_Issues__Options.pdf. Accessed May 1, 2017.

Wooldridge, J. M. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press, 2002.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

M50
MATHEMATICA
Policy Research