DEPARTMENT OF HEALTH & HUMAN SERVICES
Centers for Medicare & Medicaid Services
7500 Security Boulevard, Mail Stop S2-25-26
Baltimore, Maryland 21244-1850

CENTERS FOR MEDICARE & MEDICAID SERVICES
CENTER FOR MEDICAID & CHIP SERVICES

State Demonstrations Group

Jennifer Langer Jacobs
Director, Department of Human Services
Division of Medical Assistance and Health Services
P.O. Box 712
Trenton, NJ 08625-0712

OCT 0 1 2019

Dear Ms. Langer Jacobs:

We appreciate the efforts of you and your staff on developing the demonstration evaluation design, which is a component of the state's section 1115, titled "New Jersey FamilyCare Comprehensive Demonstration" (Project Number 11-W-00279/2). The evaluation design submitted to the Centers for Medicare & Medicaid Services (CMS) on November 24, 2017 has been found to fulfill the requirements set forth in section XIII of the Special Terms and Conditions (STC).

The evaluation design is approved for the demonstration approval period starting July 27, 2017 through June 30, 2022. Per 42 CFR 43 1.424(c), the approved evaluation design may now be posted to your state's Medicaid website.

If you have any questions, please contact your CMS project officer, Ms. Sandra Phelps. Ms. Phelps is available to answer any questions concerning your section 1115 demonstration, and her contact information is as follows:

Centers for Medicare & Medicaid Services
Center for Medicaid and Chip Services
Mail Stop: S2-25-26
7500 Security Boulevard
Baltimore, MD 21244-1850
Phone: (410) 786-1968
E-mail: Sandra.Phelps@cms.hhs.gov

We look forward to our continued partnership on the New Jersey FamilyCare Comprehensive section 1115 demonstration.

Sincerely,

Danielle Daly
Director
Division of Demonstration
Monitoring and Evaluation

Angela D. Garner
Director
Division of System Reform
Demonstrations

New Jersey FamilyCare Comprehensive Demonstration: 8/1/2017-6/30/2022

I. Evaluation Plan by Rutgers Center for State Health Policy

Background

The Special Terms and Conditions (STCs) relating to the NJ Demonstration Waiver outlines the 11 evaluation questions that are designed to examine the impact of several policy changes under the waiver on patient access to care, quality of care and costs. These policy changes include: a managed care expansion to cover long term services and supports (Questions 1 and 2); expanded income eligibility, and administrative simplifications for enrolling in managed long term services and supports (Questions 3 and 4); additional home and community-based services, and expansion of eligibility for children with intellectual and developmental disabilities and severe emotional disturbance (Questions 5, 6 and 7); cost savings from a premium assistance program for Medicaid beneficiaries who have access to employer sponsored health insurance (Question 8); expanded access and benefits for substance use disorder services (Question 9), and a three year renewal of the DSRIP program (Questions 10 and 11).

Evaluation Questions

The evaluation questions enumerated in the STCs are:

1. What is the impact of the managed care expansion on access to care, the quality, efficiency, and coordination of care, and the cost of care?
2. What is the impact of including long-term care services in the capitated managed care benefit on access to care, quality of care, and mix of care settings employed?
3. What is the impact of the hypothetical spend-down provision on the Medicaid eligibility and enrollment process? What economies or efficiencies were achieved, and if so, what were they? Was there a change in the number of individuals or on the mix of individuals qualifying for Medicaid due to this provision?
4. What is the impact of using self-attestation on the Transfer of assets look-back period of long term care and home and community based services for individuals who are at or below 100 percent of the FPL. Was there a change in the number of individuals or on the mix of individuals qualifying for Medicaid due to this provision?
5. What is the impact of providing additional home and community-based services to Medicaid and CHIP beneficiaries with serious emotional disturbance, opioid addiction, behavioral/mental health issues, or intellectual disabilities/developmental disabilities?
6. What is the impact of providing home and community-based services to expanded eligibility groups, who would otherwise have not been eligible for Medicaid or CHIP absent the demonstration?

7. What is the impact of the program to provide a safe, stable, and therapeutically supportive environment for children from age 5 up to age 21 with serious emotional disturbance who have, or who would otherwise be at risk for, institutionalization?
8. What is the impact of mandating individuals who are eligible for NJFC and have access to employee sponsored insurance into the premium assistance program; as conditional of eligibility?
9. What is the impact of providing substance use disorder services to Medicaid beneficiaries? Including paying for services rendered in an institution for mental disease (IMD)?
10. Was the DSRIP program effective in achieving the goals of better care for individuals (including access to care, quality of care, health outcomes), better health for the population, or lower cost through improvement? To what degree can improvements be attributed to the activities undertaken under DSRIP?
11. What do key stakeholders (representing covered individuals and families, advocacy groups, providers, health plans) perceive to be the strengths and weaknesses, successes and challenges of the expanded managed care program, and of the DSRIP pool? What changes would these stakeholders recommend to improve program operations and outcomes?

Managed Long-term Services and Supports

Research Questions

Q1. What is the impact of the managed care expansion on access to care, the quality, efficiency, and coordination of care, and the cost of care?

Q2. What is the impact of including long-term care services in the capitated managed care benefit on access to care, quality of care, and mix of care settings employed?

Hypothesis 1: *The managed care expansion will improve access to care, the quality, efficiency, and coordination of care, and the cost of care for the overall population in managed care.*

Hypothesis 2: Expanding Medicaid managed care to include long-term care services and supports will result in improved access to care and quality of care and reduced costs, and allow more individuals to live in their communities instead of institutions.

In New Jersey, home and community services received by the long-term care eligible population shifted from fee for service to managed care in July 2014 while the shift for nursing home residents was gradual. Members in nursing facilities at the time of enrollment were allowed to continue as fee-for-service unless they transitioned to a new level of care or moved to the community. Any new members in nursing facilities were enrolled into MLTSS. The evaluation will assess the impact of this managed care expansion to cover long-term services and supports (LTSS) over the medium and longer term, subsequent to the policy change. It will assess changes in hospitalization outcomes, preventative care rates, and measures related to spending and rebalancing

over the demonstration period compared to a baseline period, prior to the demonstration, using comparison groups to control for secular changes in such measures. The analysis will also take into account intermediate policy changes such as quality initiatives surrounding the "any willing provider" provision for nursing facility services and potential impacts on outcomes. It will examine separately specific populations of interest such as those with behavioral health (BH) conditions to examine the effect of integration of BH, physical health and LTSS under the managed long term services and supports (MLTSS).

Outcome Measures

Claims-based: Avoidable hospitalizations and ED visits; 30-day readmission rates; rates of follow up care after any hospitalization and after mental health hospitalization; overall rates of hospitalization and ED visits; avoidable inpatient and ED hospital spending; HbA1c testing; diabetic eye exam; LDL Screening; dental utilization; share of first time LTSS users receiving HCBS (rather than institutional services); share of all LTSS beneficiaries using HCBS; per capita LTSS spending; HCBS share of total LTSS spending.

HEDIS and CAHPS®: Quality measures related to preventive care, behavioral health, chronic conditions, and consumer satisfaction.

Metrics reported by MCOs, EQROs, State Government, and other partners: While we do not possess the data utilized for creating these metrics (as we do the claims data), we will review reports by such entities, such as the MLTSS Quality Metrics reported by managed care organizations (MCOs), state departments, and external quality review organizations (EQROs). We will also review the National Core Indicators—Aging and Disability reports. If furnished reports contain metrics that are relevant for measuring access to care and quality of care and for exhibiting trends over time, we will include them as context in our reporting. In past evaluation reports, we presented data on assessment timeliness, critical incidents and appeals, complaints and grievances, assessments of care plans and the timeliness of service onset. We also presented the current status of former waiver enrollees, which showed that they have been able to remain in community settings rather than transitioning to nursing homes. With respect to the NCI-AD, we examined and reported differences in participant demographics and outcomes between the following groups: MLTSS enrollees in New Jersey with MLTSS enrollees in other participating states; MLTSS enrollees in New Jersey compared with other LTSS programs in New Jersey; and MLTSS enrollees among different MCOs in New Jersey. The frequency of data reporting varies for these sources—some are monthly, some quarterly, some semiannually and others annually.

Stakeholder feedback: We will conduct approximately 20 interviews with MLTSS stakeholders. Stakeholders are defined as representatives of organizations that serve a client group also served by MLTSS, and we anticipate that they will include consumer advocates, provider associations, community partner agencies (such as County Welfare

Agencies, Area Agencies on Aging, Centers for Independent Living, and local emergency responders), managed care organizations, and state officials. Potential interviewees will be identified based on membership in the MLTSS Steering Committee that has advised state officials before and after MLTSS implementation, recommendation of Steering Committee members, representatives who have contacted the Center for State Health Policy (CSHP) based on prior waiver evaluation work, or additional organizations identified by CSHP as serving a relevant population. At a minimum, we will invite for interviews representatives that serve the different waiver populations as defined prior to MLTSS, including older adults, younger adults and children with disabilities (physical, developmental, and traumatic brain injury), and children and adults with HIV/AIDS. We will ask questions about their views on the impact of MLTSS on the population groups with whom they work with respect to service adequacy, care management, continuity of care, and access to services in community settings, as well as how MLTSS has evolved over time. We will also ask about impacts on providers and other community partners, such as Area Agencies on Aging and Centers for Independent Living.

Administrative Simplifications in Eligibility and Enrollment

Research Questions

Q3. What is the impact of the hypothetical spend-down provision on the Medicaid eligibility and enrollment process? What economies or efficiencies were achieved, and if so, what were they? Was there a change in the number of individuals or on the mix of individuals qualifying for Medicaid due to this provision?

Q4. What is the impact of using self-attestation on the transfer of assets look-back period of long term care and home and community based services for individuals who are at or below 100 percent of the FPL. Was there a change in the number of individuals or on the mix of individuals qualifying for Medicaid due to this provision?

Hypothesis 3: Utilizing Qualified Income Trusts will allow more individuals to qualify for Medicaid and will increase the number of Medicaid long-term care recipients in community settings.

Hypothesis 4: Eliminating the look back period at time of application for transfer of assets for applicants or beneficiaries seeking long term services and supports whose income is at or below 100% of the FPL will simplify Medicaid eligibility and enrollment processes without compromising program integrity.

Qualified Income Trusts (QITs), which are the mechanism through which enrollees qualify for long-term care services if their income exceeds eligibility limits, effectively create a new eligibility pathway for long-term care services in home and community settings. QITs allow clinically eligible individuals whose monthly income is above 300% of the Supplemental Security Income rate to have excess income disregarded in determining Medicaid eligibility. Income above the threshold is deposited in a separate

bank account which is dedicated exclusively to Medicaid-approved uses. The introduction of the QIT mechanism required discontinuing the Medically Needy program which reduced the resource limits for eligibility for nursing home care to community levels.

Also under the initial demonstration and continuing in the renewal, individuals with income at or below 100% of the Federal Poverty Level (FPL) applying for institutional or home and community-based services are permitted to self-attest that they have made no disqualifying asset transfers during the past five years. This procedure is intended to expedite eligibility approvals for very low-income applicants by eliminating the need for the time intensive five-year lookback process.

The evaluation will examine outcome measures related to the implementation of these administrative simplifications. We will examine changes in the mix and characteristics of individuals qualifying for Medicaid LTSS by setting of care in the pre and post-policy periods. Contingent on the availability of published reports or administrative data collected by the State, we will examine the extent to which QIT use varies by long-term care setting (nursing facility (NF), assisted living (AL), home and community-based settings (HCBS)) and characteristics of QIT users.

Outcome Measures

Claims-based

QIT: Proportion of LTSS beneficiaries in NF, AL, HCBS

Audit data from Bureau of Quality Control

Self-attestation: Error rate on audited self-attestations

Published reports and communications with State representatives

QITs: Number of submitted, eligible, and approved QITs each quarter overall and by setting of care; Proportion of QIT users who are in the community; Volume of QIT use by county.

Self-attestation: Number of self-attestations received each quarter overall and by county, setting of care, and MCO

Targeted Home and Community-Based Services for Children and Youth

Research Questions

Q5. What is the impact of providing additional home and community-based services to Medicaid and CHIP beneficiaries with serious emotional disturbance, opioid addiction[1], behavioral/mental health issues, or intellectual disabilities/developmental disabilities?

[1] Examination of waiver polices affecting beneficiaries with opioid addiction will be conducted under research question 9 which is addressed in a standalone evaluation plan.

Q7. What is the impact of the program to provide a safe, stable, and therapeutically supportive environment for children from age 5 up to age 21 with serious emotional disturbance who have, or who would otherwise be at risk for, institutionalization?

Hypothesis 5: Providing home and community-based services to Medicaid and CHIP beneficiaries and others with serious emotional disturbance or intellectual disabilities/developmental disabilities with and without co-occurring mental illness will lead to better care outcomes including those relating to ambulatory care.

Hypothesis 7: Providing home and community-based services to Medicaid and CHIP beneficiaries and others with serious emotional disturbance who have, or who would otherwise be at risk for, institutionalization will reduce avoidable utilization.

The Children's Support Services Program (CSSP) absorbs the pilot programs for children with serious emotional disturbance (SED) and children with intellectual/developmental disabilities and a co-occurring mental health diagnosis (ID-DD/MI) administered by the Division of Children and Families' Children's System of Care (DCF-CSOC). It also covers ID-DD children without a co-occurring mental health diagnosis. Under the CSSP, eligible children can receive targeted home and community-based services and supports and/or behavioral health services which promote their success and stability in a community setting. The pilot for children with Autism Spectrum Disorder (ASD) will continue under the demonstration until approval of a State Plan Amendment which will incorporate the services into the NJ Medicaid State Plan.

The Supports Program was initiated under the 2012-2017 Waiver to provide a basic level of support services to Medicaid adults with intellectual disabilities/developmental disabilities who live with family members or in other unlicensed settings in the community. This program continues under the Waiver renewal. The Community Care Waiver, formerly excluded from the 1115 Waiver, came under 1115 authority as the Community Care Program (CCP). The CCP provides services and supports to Medicaid adults meeting the ICF-ID level of care requirements who reside in the community.

The evaluation will characterize the populations and assess volume and array of service use in the CSSP, Supports, and CCP. It will assess relevant outcome measures over the pre- and post-policy period for individuals receiving these additional services to examine potential effects of this policy change. We will construct comparison groups, for instance, matching youth receiving waiver services with Medicaid youth having ID-DD or SED, but uninvolved with DCF-CSOC. We will examine the appropriateness of such comparison groups for isolating the policy impact by comparing demographic and clinical characteristics of the intervention and comparison groups and also qualitatively, through discussions with state policymakers. We will also look at outcomes among young adults who had and did not have services under DCF-CSOC waiver programs to

determine the extent to which the waiver services supported the transition to adulthood for these youth.

Outcome Measures

Claims-based

ASD: overall inpatient hospitalizations; avoidable hospitalizations; ED visits; avoidable ED visits; 30-day readmissions; stays in out-of-home care settings; well-child visits; avoidable and overall hospital spending per beneficiary.

ID-DD: overall inpatient hospitalizations and length of stay; avoidable hospitalizations; ED visits; avoidable ED visits; 30-day readmissions; stays in out-of-home care settings; well-child visits; avoidable and overall hospital spending per beneficiary.

ID-DD/MI: overall inpatient hospitalizations and length of stay; avoidable hospitalizations; ED visits; avoidable ED visits; 30-day readmissions; inpatient stays for mental health conditions, stays in out-of-home care settings; well-child visits.

SED at-risk: stays in out-of-home care settings

SED 217-like: overall inpatient hospitalizations and length of stay; avoidable hospitalizations; ED visits; avoidable ED visits; 30-day readmissions; inpatient stays for mental health conditions, stays in out-of-home care settings; well-child visits.

Supports: Rates of Hemoglobin A1C Testing, Pneumococcal Vaccination, diabetic eye exam, follow up after hospitalization for mental illness; IDD specific preventable hospitalizations (e.g., epilepsy, Gastro-esophageal reflux disease).

CCP: Rates of Hemoglobin A1C Testing, Pneumococcal Vaccination, diabetic eye exam, follow up after hospitalization for mental illness; IDD specific preventable hospitalizations (e.g., epilepsy, Gastro-esophageal reflux disease).

DCF-CSOC Reported Quality Metrics

ID-DD, ID-DD/MI, and SED: Improvement in Child and Adolescent Needs and Strength composite rating; Services delivered in accordance with the approved plan of care; CSOC verification that providers of waiver services continually meet required qualified status; Percentage of Unusual Incident Reports submitted involving waiver participants

Eligibility Expansions for Populations in Need of Home and Community-Based Services

Research Question

Q6. What is the impact of providing home and community-based services to expanded eligibility groups, who would otherwise have not been eligible for Medicaid or CHIP absent the demonstration?

Hypothesis 6: Providing home and community-based services to expanded eligibility groups, who would otherwise have not been eligible for Medicaid or CHIP absent the demonstration will lead to improvements in preventive care and avoidable utilization.

The CSSP-ID/DD allows for expanded Medicaid eligibility for children meeting functional criteria and having a plan of care with CSOC's Care Management Organization. Children with income up to 300% FBR receive State Plan services and waiver home and community-based services. Eligibility for the Supports Program also allows individuals up to 300% FBR to receive Medicaid State Plan and waiver home and community-based services.

The income eligibility expansions authorized under the 2012-2017 demonstration for children with SED and the adoption of Qualified Income Trusts for higher-income individuals in need of long-term care services continue under the waiver renewal.

The evaluation will identify individuals in the data who, absent the demonstration, would not have been eligible for Medicaid. It will characterize the volume and patterns of service use for the expansion populations and assess relevant outcome measures for individuals receiving these additional services to examine potential effects of this policy change. When feasible, we will construct appropriate comparison groups to help isolate the policy impact, and in the absence of such appropriate controls, will investigate differences in beneficiary characteristics and service use between those with favorable versus unfavorable outcomes.

Due to the absence of baseline data for these populations (since prior to the policy change they were not Medicaid-eligible and hence would not show up in our claims data), we will conduct trend analyses of outcomes over time only after policy implementation.

Outcome Measures

Claims-based

CSSP: overall inpatient hospitalizations and length of stay; avoidable hospitalizations; ED visits; avoidable ED visits; 30-day readmissions; inpatient stays for mental health conditions, stays in out-of-home care settings; Well-child visits.

Supports: Rates of Hemoglobin A1C Testing, Pneumococcal Vaccination, diabetic eye exam, follow up after hospitalization for mental illness; IDD specific preventable hospitalizations (e.g., epilepsy, Gastro-esophageal reflux disease).

MLTSS: Avoidable hospitalizations and ED visits; 30-day hospital-wide and pneumonia readmission rates; rates of follow up care after hospitalization; overall rates of hospitalization and ED visits; HbA1c Testing; diabetic eye exam; LDL Screening

Premium Support Program

Research Question

Q8. What is the impact of mandating individuals who are eligible for NJFC and have access to employee sponsored insurance into the premium assistance program; as conditional of eligibility?

Hypothesis 8: Mandating individuals who have access to employee sponsored insurance into the premium assistance program will cost the State at least 5% less than providing individuals coverage in NJFC.

The Premium Support Program (PSP) will provide premium reimbursement to NJFC-eligible individuals with access to health insurance through an employer if such reimbursement is determined to be more cost-effective than NJFC enrollment. If the employer-sponsored insurance (ESI) plan is not equivalent to at least the NJFC Plan D service package, then wraparound NJFC services are provided. In addition, NJFC-eligible individuals enrolled in ESI through the PSP have their out-of-pocket costs capped, with NJFC covering any payments incurred in excess of 5% of gross income.

We will use data provided by DMAHS to calculate the actual net cost savings due to a Medicaid beneficiary (and any eligible dependents) enrolling in the premium support program. This will be calculated using costs incurred by Medicaid for a beneficiary enrolled in the PSP (premium reimbursement +wraparound benefit +cost sharing above 5% cap) less the cost incurred if this person were enrolled in NJFC instead of the PSP.

Outcome Measures

DMAHS PSP Net Savings to NJ Data Report: Per-member per-month net savings due to PSP.

Provision of substance use disorder services

Research Question

Q9. What is the impact of providing substance use disorder services to Medicaid beneficiaries? Including paying for services rendered in an institution for mental disease (IMD)?

The SUD initiative is addressed in a standalone evaluation plan that will be provided in a separate document

The Delivery System Reform Incentive Payment Program

Research Question

Q10. Was the DSRIP program effective in achieving the goals of better care for individuals (including access to care, quality of care, health outcomes), better health for the population, or lower cost through improvement? To what degree can improvements be attributed to the activities undertaken under DSRIP?

Q11. What do key stakeholders (covered individuals and families, advocacy groups, providers, health plans) perceive to be the strengths and weaknesses, successes and challenges of the expanded managed care program, and of the DSRIP pool? What changes would these stakeholders recommend to improve program operations and outcomes?

See Section II for the detailed evaluation plan related to the DSRIP.

Measure Definitions

The table below provides details on the proposed measures for evaluation of Research Questions 1-8.

ANALYTIC STRATEGY

The component of the evaluation examining research questions 1-8 (we have separate analytic strategies for the DSRIP and SUD demonstration) will utilize both quantitative as well as qualitative analysis. The quantitative component will involve analysis of Medicaid claims/encounter data and aggregated or summary statistics from secondary sources. The claims data provides information on patient, provider and geographic characteristics, and we will adjust for such factors while examining the policy effects on our outcomes of interest. We will not have such information for secondary metrics that we may use to provide context but will calculate statistical significance of annual trends wherever possible.

The qualitative component will be key informant interviews that will capture stakeholder perceptions relating to program implementation, potential, and perceived impacts.

Quantitative Analysis

This description, specifically the multivariate statistical analysis, is mostly relevant to the claims data analysis where it is possible to adjust for patient and provider characteristics and examine trends over time. Depending on the frequency at which summarized statistics from secondary sources are available, we will construct trends and examine for statistical differences.

Measure Descriptions and Crosswalk to Hypotheses for Research Questions 1-8

Measure	Source	NQF	Description/Numerator	Denominator (f)	Hypotheses
Source: Medicaid Claims and Encounter Data					
Inpatient (IP) hospitalizations			Inpatient stays at general acute care hospitals	(g)	1, 2, 5, 6, 7
Inpatient days			Number of days for inpatient stays at general acute care hospitals	(g)	5, 6, 7
Emergency department (ED) visits			Treat-and-release emergency department visits	(g)	1, 2, 5, 6, 7
Overall hospital spending (IP+ED)			Payments on facility claims for inpatient and treat-and-release ED visits	(g)	5
Avoidable hospitalizations	AHRQ		Prevention Quality Indicator (PQI) #90 and Pediatric Quality Indicator (PDI) #90 are potentially avoidable hospitalizations for ambulatory care sensitive conditions that reflect issues of access to, and quality of, ambulatory care in a given geographic area.	Medicaid recipients age 6-17 (PDI #90); Medicaid recipients age 18 and older (PQI #90)	1, 2, 5, 6, 7
Avoidable inpatient hospitalization costs			Payments on facility claims for avoidable inpatient visits	(g)	1, 2
Avoidable ED visits	(a)		Treat-and-release emergency department visits that are: -Non-emergent -Emergent/primary care treatable -Emergent, ED care needed - preventable/avoidable -Emergent, ED care needed - not preventable/avoidable	(g)	1, 2, 5, 6, 7
Avoidable ED visit costs			Payments on claims for avoidable treat-and-release ED visits	(g)	1, 2
Overall avoidable hospital spending (IP+ED)			Payments on facility claims for avoidable inpatient and avoidable treat-and-release ED visits	(g)	5

Measure Descriptions and Crosswalk to Hypotheses for Research Questions 1-8

Measure	Source	NQF	Description/Numerator	Denominator (f)	Hypotheses
Inpatient stays for mental health conditions			Hospitalizations with a primary diagnosis of mental illness	Medicaid recipients ages 6 and older	5, 6, 7
Follow-up after hospitalization			Ambulatory visit 7 or 14 days after discharge from an inpatient stay	Hospital discharges to a home/community setting; excludes patients discharged against medical advice.	1, 2, 6
Follow-up after mental illness hospitalization	NCQA	576	Percentage of discharges for Medicaid recipients hospitalized for treatment of selected mental illness diagnoses who had a follow-up visit with a mental health practitioner within 7 and 30 days of discharge	Hospital discharges to a home/community setting with a primary diagnosis of mental illness for Medicaid recipients age 6 and older	1, 2, 5, 6
HbA1c testing	NCQA	57	Percentage of adult patients receiving one or more A1c test(s) per year	Medicaid recipients ages 18-75 with diabetes	1, 2, 5, 6
Diabetic Eye Exam	NCQA	55	Percentage of adult patients who received an eye screening for diabetic retinal disease during the measurement year.	Medicaid recipients ages 18-75 with diabetes	1, 2, 5, 6
LDL screening	NCQA	63	Percentage of adult patients receiving one or more LDL-C tests per year	Medicaid recipients ages 18-75 with diabetes	1, 2, 6
Annual dental visit	NCQA	1388	Percentage of Medicaid recipients who had at least one dental visit during the measurement year	Modified from measure steward's age specifications of 2-20 years to apply to Medicaid recipients of all ages.	1, 2
Frequency of stays in out-of-home care settings			Stays in an accredited residential treatment facility for youth	Medicaid recipients up to age 20	5, 6, 7

Measure Descriptions and Crosswalk to Hypotheses for Research Questions 1-8

Measure	Source	NQF	Description/Numerator	Denominator (f)	Hypotheses
Well-Child Visits in the Third, Fourth, Fifth, and Sixth Years of Life	NCQA	1516	Percentage of Medicaid recipients who received one or more well-child visits with a PCP during the measurement year.	Medicaid recipients 3 to 6 years of age	5, 6, 7
Pneumococcal Vaccination for Older Adults	NCQA (b)		Percentage of Medicaid recipients who have received the recommended series of pneumococcal vaccines	Medicaid recipients age 65 and older	5, 6
Hospitalization for epilepsy	(c)		Rate of potentially avoidable hospitalizations for ambulatory care sensitive conditions applicable to persons with an intellectual disabilities that reflect issues of access to, and quality of, ambulatory care in a given geographic area.	Medicaid recipients with intellectual/ developmental disabilities	5, 6
Hospitalization for GERD	(c)				5, 6
Hospitalization for constipation	(c)				5, 6
Hospitalization for schizophrenic disorders	(c)				5, 6
30-day hospital-wide all-cause readmissions	CMS	1789	Percentage of discharges followed by an unplanned readmission to any acute care hospital within 30 days of discharge.	Hospital discharges to a home/community setting for Medicaid recipients age 18 and older; excludes patients discharged against medical advice	1, 2, 5, 6, 7
30-day pneumonia readmission	CMS	506	Percentage of discharges followed by an unplanned readmission to any acute care hospital within 30 days of discharge from a hospital.	Hospital discharges to a home/community setting for Medicaid recipients age 18 and older following a hospitalization with a primary diagnosis of pneumonia; excludes patients discharged against medical advice	6

Measure Descriptions and Crosswalk to Hypotheses for Research Questions 1-8

Measure	Source	NQF	Description/Numerator	Denominator (f)	Hypotheses
LTSS spending			Payments on claims for long-term services and supports	All long-term care Medicaid recipients	2
Share of first-time LTSS users receiving HCBS	(d)		Medicaid recipients entering MLTSS who receive services in a home or community-based setting in their first month of receiving LTSS.	Medicaid recipients entering MLTSS	2
Share of all LTSS beneficiaries using HCBS			Medicaid recipients in MLTSS receiving services in a home or community-based setting for the majority of their program enrollment	Medicaid recipients in MLTSS	2
HCBS share of total LTSS spending			Spending for home and community-based long-term care services	Spending for all long-term care services	2
LTSS beneficiaries by setting of care			Proportion of all long-term care Medicaid recipients in nursing facilities, assisted living, and living at home.	All long-term care Medicaid recipients	3
Source: Secondary Data (e)					
HEDIS quality metrics for NJ Medicaid MCOs	NCQA		Performance of Medicaid managed care organizations on metrics related to quality of preventive care, treatment of chronic conditions, and behavioral health care. Example metrics are: -Childhood vaccinations rates -Rates of follow-up after mental illness hospitalizations -Rates of blood pressure control	(h)	1
CAHPS survey results for NJ Medicaid MCOs	NCQA		Consumer satisfaction with care provision under managed care. Example metrics are perceptions around: -Getting care quickly -How well doctors communicate -Personal doctor informed about care from other providers	(h)	1, 2 (i)

Measure Descriptions and Crosswalk to Hypotheses for Research Questions 1-8

Measure	Source	NQF	Description/Numerator	Denominator (f)	Hypotheses
Metrics reported by MCOs, EQROs, and State Government			Quality metrics related to MLTSS reported by MCOs and data on MLTSS progress reported by the State to stakeholders. Example metrics are: -Assessment timeliness -Assessment of care plans -Status of former 1915(c) waiver enrollees	(h)	2
National Core Indicators - Aging and Disability	NASUAD and HSRI		Survey data for long-term care populations assessing receipt of services, satisfaction with services, and quality of life. Example metrics are: -Whether assistance received meets needs and goals -Whether people feel in control over the life -Utilization of health services	(h)	2
Use of Qualified Income Trusts (QITs)			Number of submitted, eligible, and approved QITs; Proportion of QIT users who are in the community; Volume of QITs use by county	Number of QITs	3
Use of self-attestations			Number of self-attestations received by State overall and by setting of care.	Number of self-attestations	4
Error rate on audited self-attestations			Proportion of audited self-attestations having a transfer of assets in the past five years	Number of sampled and audited self-attestations	4
Division of Children and Families - Children's System of Care (CSOC) Quality Metrics			Quality metrics from the CSOC Quality Strategy. Example metrics are: -Improvement in child and adolescent needs and strength composite rating -Services delivered in accordance with plan of care -Percentage of unusual incident reports	(h)	5, 7

Measure Descriptions and Crosswalk to Hypotheses for Research Questions 1-8

Measure	Source	NQF	Description/Numerator	Denominator (f)	Hypotheses
			submitted involving waiver participants		
Cost savings for Premium Support Program (PSP)			Net savings calculated as the difference between costs to Medicaid for NJ FamilyCare enrollment and costs for PSP.	PSP member months	8

AHRQ = Agency for Healthcare Research Quality; NCQA = National Committee for Quality Assurance; CMS = Centers for Medicare & Medicaid Services; LTSS= Long-term Services and Supports; MCO=Managed Care Organization; NASUAD = National Association of States United for Aging and Disability; HSRI=Human Services Research Institute

(a) https://wagner.nyu.edu/faculty/billings/nyued-background

(b) This is an electronic clinical data system measure introduced in HEDIS 2018 which we will calculate using Medicaid claims.

(c) Balogh, R. S., Ouellette-Kuntz, H., Brownell, M.,& Colantonio, A. (2011). Ambulatory care sensitive conditions in persons with an intellectual disability - Development of a consensus. J of Applied Research in Intellectual Disabilities, 24, 150-158.

(d) Long-term Care Scorecard, http://www.longtermscorecard.org/~/media/Microsite/Files/2017/2_RankingMethodology_June12_v2.pdf.

(e) Review and analysis of all secondary data is contingent upon availability and completeness of data received from the State.

(f) General inclusion or exclusion criteria (if any) for the denominator are noted here. Any other inclusion or exclusion criteria in measure specifications will also be followed (e.g. history of certain conditions, length of enrollment, etc.). Measures will also be calculated for subpopulations relevant to each hypothesis. See description of target and comparison populations in Analytic Strategy section.

(g) No denominator inclusion or exclusion criteria for this measure.

(h) Measures are not independently calculated. Numerator and denominator criteria are set by the agency collecting and calculating these measures.

(i) CAHPS data can be used to address hypothesis 2 if reported specifically for the managed care subpopulation in MLTSS.

We first describe the general aspects of different statistical models that are applicable to multiple research questions and the related hypotheses. We also provide information on the data used for the quantitative analysis.

Next we have specific subsections providing further details on analysis pertaining to specific research questions such as pre-post periods, statistical modeling approach or comparison groups when relevant.

Data: Depending on the particular analysis, we will utilize Medicaid claims and managed care encounter data over the period January 2011 to June 2022 utilizing a minimum six month runout period. The State has estimated that the majority of FFS and managed care claims are received within six months of the date of service, and this lag efficiently balances data completeness with the timely completion of analyses. Monthly extracts are received and used to build static analytic claims files. Our analytic files are validated against a real-time database query from DMAHS on total payment amounts, total number of claims, and recipient eligibility counts for a specified period and differ by <1%. Additionally, constructed population indicators (e.g. nursing facility residents, children enrolled in DCF-CSOC waivers, etc.) are always benchmarked against State figures for these same populations when available.

New Jersey managed care plans must submit all services provided to MLTSS recipients to the State. The accuracy and completeness of provider payment amounts reported on these encounter claims is assured through a number of validation checks. First, service encounters are reviewed for accuracy by New Jersey's fiscal agent before being considered final. The State implements liquidated damages on its health plans for excessive duplicate encounters and excessive denials. Further, accurate payment reporting processes are ensured by the requirement that after a defined period of time the total dollar value of encounters accepted by the State's fiscal agent must also equal 98 percent of the medical cost submitted by the plans in their financial statements.

Our claims database is constructed with all the updates, voids, and adjustments to costs available from the State at the point of construction with no month having less than six month runout period. This structure was decided in consultation with the State to balance data completeness with the timely completion of evaluation analyses.

Medicare claims will not be available for this evaluation. Utilization is available for fee-for-service dually eligible beneficiaries in our Medicaid claims database. Utilization by managed care duals is present in our Medicaid claims database if there is a Medicaid liability for the encounter. Such liability arises when Medicaid covers the co-insurance and any cost difference between the provider charges and Medicare reimbursement so that dual beneficiaries are not billed for medically necessary services. In a limited number of situations where there is no Medicaid liability at all for the encounter, the presence of the utilization in our database is dependent on MCO reporting protocols.

Although we expect any undercount of utilization, especially for hospitalization outcomes, to be minimal, our analytic strategy (described below) utilizes difference-in-differences to evaluate the impact of MLTSS which further mitigates data incompleteness issues. We select our control group so as to achieve balance on a number of covariates that may affect outcomes. Similarly we will balance our MLTSS and comparison group on dual eligibility status so that both are similarly affected by any residual outcome measurement issues related to their dual status. All analyses will include a control for dual eligibility status.

Only spending by Medicaid will be counted in outcome measures related to costs consistent with our focus on Medicaid spending.

Pre- and post-implementation period: Analysis of Medicaid claims data will entail examining changes in the levels and trends of the selected metrics (relating to each hypothesis) subsequent to the policy implementation. Measuring differences in these outcomes between time periods before and after the implementation of the program/policy change will identify the program effect. During such identification we will incorporate wherever feasible, trends in comparison groups to account for secular changes unrelated to the policy effects (see greater discussion of this in the difference-in-differences section below). For policies in the renewal demonstration period that are related to those in the initial demonstration, we will assess potential changes in trends over three distinct periods. These include the baseline period for the first evaluation: January 1, 2011-September 30, 2012; the first demonstration period: Oct 1, 2012–July 31st, 2017; and the second demonstration period: August 1, 2017-June 30, 2022. The statistical model will account for these three distinct periods by incorporating indicator variables for specific years or rounds of demonstration. This will allow estimation of changes in outcomes during the first demonstration period from policy changes, and additional changes in outcomes during the second demonstration period from continuation of those policy changes. For new policies during the second demonstration period, such as those relating to SUD services, we will examine a baseline period prior to the time of policy implementation and examine changes in outcomes between the baseline and the post-implementation period.

Difference-in-Differences Estimation: For estimating the policy effect, the evaluation will utilize a difference-in-differences (DD) estimation technique when it is possible to define appropriate comparison groups for the study population. DD modeling identifies the impact of the policy change by comparing the trend in outcomes for the program eligible/targeted (intervention) population from the pre- to the post-implementation period to that of a comparison group which is otherwise similar, but not subject to the policy effect. Such an estimation strategy is able to identify changes in outcomes that are due to program impact and distinct from secular trends. It accounts for the effect of unobserved factors, as long as their impact on one of the groups relative to the other

does not change over time. This last assumption is tested by examining whether trends in outcomes prior to policy implementation (pre-trends) for the intervention and comparison group are parallel to each other. This is described in detail in the next section.

Examining validity of DD estimates: The crucial assumption relating to the DD approach is there are no unmeasured factors whose effect on the intervention group relative to the comparison groups changes over time. This may not always be fulfilled. In that case, the unobserved factors may result in the two groups having differential pre-policy trends (pre-trends), and the computed effect size will need to adjust for this difference in pre-trends. Accordingly, we will test to see whether there existed statistically significant differences in trends between the intervention and comparison group prior to policy implementation. If this difference is in the same direction as the DD estimate and of comparable magnitude that would imply that the DD model may be overestimating the effect. Accordingly our estimated regression coefficient providing the policy effect will be adjusted for these differential pre-trends based on well-established methods in peer-reviewed academic publications.[2]

Segmented Regression Analysis: While we will develop comparison groups wherever feasible in our evaluation analyses to facilitate separation of program impact from secular trends, it may not be always possible to have suitable comparison groups. In those cases we will use Segmented Regression Analysis. Such a model assumes that the policy effect may lead to a change in level, and also a change in the existing time trend of the metric measuring quality or any other relevant outcome of interest. The regression analysis is able to measure this change in trend or level. Potential confounding may arise in the rare circumstances when factors that determine our outcomes of interest change at exactly the same time as the policy implementation. However, our multivariate analysis adjusting for patient, provider and geographic factors are expected to mitigate such effects. As shown in our previous evaluation work,[3] this approach also allows us to model the effect of separate policy changes at other points of time, and separate those effects from our policy of interest.

[2] Harman, J. S., Hall, A. G., Lemak, C. H., & Duncan, R. P. (2014). Do provider service networks result in lower expenditures compared with HMOs or primary care case management in Florida's Medicaid program? Health Serv Res, 49(3), 858-877. PMCID: PMC4231575

[3] Chakravarty, S., Lloyd, K., Farnham J., Brownlee, S., & DeLia D. (2017). Examining the Effect of the NJ Comprehensive Waiver on Access to Care, Quality, and Cost of Care: Draft Final Evaluation Report. New Brunswick, New Jersey: Rutgers Center for State Health Policy. Available at: http://www.cshp.rutgers.edu/publications/examining-the-effect-of-the-nj-comprehensive-waiver-on-access-to-care-quality-and-cost-of-care-draft-final-evaluation-report.

Adjusting for Patient, Provider and Geographic Factors: Our multivariate analysis will control for patient characteristics that may affect outcomes. These include beneficiary demographics, Medicaid eligibility category, health history (including chronic illness and behavioral health co-morbidities), chronic disability payment score, and any other information relevant to the policy of interest. We will incorporate hospital fixed effects (to account for time-invariant differences across hospitals) for inpatient quality-based measures and zip code fixed effects (to account for time-invariant measures across geographic locations) for measures reflecting ambulatory care. We will utilize when required, statistical matching techniques such as "Mahalanobis matching" or propensity score matching to create comparison cohorts of patients unaffected by policy changes for patients subject to policy effects.

Dose Response: Wherever applicable and relevant we will examine whether there is a "dose-response" relationship. Findings of a higher response when the "dose" of a policy change will strengthen causal inferences.

Methodological Limitations: As mentioned above, it may sometimes not be possible to generate an appropriate comparison group if the policy universally impacts a broad category of beneficiaries, for instance, individuals with a particular behavioral health condition. In addition, sometimes data relating to a pre-policy baseline period are not available, if the beneficiaries are newly Medicaid-eligible, or reported data is collected only after policy implementation. In that case we will assess time trends in the post-policy period and assess changes in outcomes over time. Our ability to calculate metrics and determine accurate policy effects may be limited by accuracy and availability of program status codes and relevant data.

We next provide information on specific aspects of the statistical modeling that are distinct to the individual research questions and for testing related hypotheses.

Research Questions 1 & 2 relating to MLTSS: In New Jersey, all LTSS eligible individuals living in the community, and receiving home and community based services (HCBS) shifted from fee-for-service to managed care for their LTSS in July 2014. Individuals residing in the nursing facilities shifted more gradually to managed care and the enrollment trigger was transitioning to a new facility or the community. Because of such differences in the managed care enrollment process, and also in the extent of disability between individuals receiving HCBS and those in the NFs, we will separately examine the effect of MLTSS on these two populations.

For the population receiving HCBS, the DD analysis will compare changes in outcomes from the pre (January 2011-June 2014) to the post- period (July 2014-June 2022) for this treatment group relative to a comparison group of individuals selected from the Medicaid 'aged, blind, disabled' (ABD) eligibility category who do not receive such LTSS

services. This comparison group is utilized to account for trends in outcomes unrelated to the MLTSS policy implementation.

Statistical methods for incorporating comparison group in DD analysis: We will use propensity score analysis while selecting Medicaid beneficiaries categorically eligible as ABD as comparison individuals. Such a method takes into account patient characteristics determining evaluation outcomes that may also determine the likelihood of receiving HCBS. An initial logistic regression models the likelihood of receiving HCBS in the sample of community-based Medicaid beneficiaries (that include our treatment group and the ABD group of beneficiaries) as a function of characteristics that determine the likelihood of receiving HCBS. Such variables may include age, sex, behavioral health, dual eligible status, chronic disability payment score and enrollment history. The predicted probabilities from this model will be used to weigh observations in the comparison group that are above a threshold probability level. Incorporating such propensity score reweighting (Nichols, A, 2007, 2008)[4] will generate an optimal comparison group for the difference-in-differences analysis that is similar to the intervention group.

NF residents: For the NF residents, we will utilize similar methods to generate a comparison group using propensity score modeling. However, we will also utilize additional analytic techniques since the comparison categorically eligible ABD group are community-dwelling and may differ in unobserved ways from the NF residents in terms of disability and health. Accordingly, we will examine changes in outcomes of NF individuals as they transition from FFS to managed care. While we will not be able to use the traditional interrupted time series design[5] since the transition occurs for different individuals at different points of time, the proposed analytic technique utilizes a similar identification strategy. Changes in outcomes of individuals that are contemporaneous with exposure to the policy (when they transition to FFS to managed care) will be estimated through regression analysis. We will also conduct sensitivity analysis through a falsification test that estimates a placebo model by excluding data after 2014 and falsely assuming that the policy change was implemented in 2013. Based on methods previously used by the evaluation team[6], this examines whether there were any

[4] Nichols, A. 2007. Causal inference with observational data. Stata Journal 7: 507–541; Nichols, A. 2008. Erratum and discussion of propensity–score reweighting. The Stata Journal. 2008. Volume 8 Number 4: pp. 532-539.

[5] Wagner AK, SB Soumerai, F Zhang, and D Ross-Degnan. 2002. "Segmented Regression Analysis of Interrupted Time Series Studies in Medication Use Research." *Journal of Clinical Pharmacy and Therapeutics* 27 (4): 299–309.

[6] Cantor, J.C., Monheit, A.C., DeLia, D. and Lloyd, K. (2012). Early impact of the affordable care act on health insurance coverage of young adults. Health Serv Res, 47(5), 1773-90.

statistically significant changes in outcomes, one year prior to the change in financing from FFS to managed care.

Research Questions 3 & 4 relating to Administrative Simplifications: Suitable comparison populations are not available among Medicaid beneficiaries and will not be used in evaluating the hypotheses for these research questions.

Research Question 6 relating to Eligibility Expansion for populations receiving HCBS: The policy change of expanded Medicaid eligibility results in a study population that is a newly enrolled group of Medicaid beneficiaries. We will isolate a cohort of these newly eligible beneficiaries to the extent possible in the claims data. However, being limited to Medicaid data, we cannot identify healthcare utilization for this study population during their pre-period. We will examine their trends in health outcomes subsequent to Medicaid enrollment that will shed light on the long term impact of the policy.

Research Questions 5 and 7 relating to HCBS services for Medicaid and CHIP beneficiaries: We will utilize a DD strategy utilizing comparison groups for each of the three study populations of children: with ASD, ID-DD(/MI) and SED receiving home and community services. Comparison groups will be Medicaid/CHIP beneficiaries identified in the Medicaid claims having similar diagnosis and demographics, but not receiving waiver services. The DD estimate will shed light on the policy effect by estimating the pre-post change in outcomes for the study population relative to the comparison population. As discussed above, we will examine whether pre-trends are parallel and if not, will account for such trends using methods discussed above.

Research Question 8 relating to the Premium Support Program: We will utilize comparison estimates that indicate costs if the beneficiaries in the Premium Support Program were to instead be covered under NJ FamilyCare.

Research Question 9 relating to the OUD/SUD initiative: This is a standalone evaluation plan that will be provided in a separate document.

Research Questions 10 and 11 relating to DSRIP: Please see the DSRIP section for potential comparison groups in DD analysis, alternative strategies including interrupted time series modelling and sensitivity analysis including falsification tests, and checking pre-trend parallel assumption.

Qualitative Analysis

Qualitative analysis regarding the DSRIP program appears later. Regarding our MLTSS interviews, interviewers will use a semi-structured guide containing key questions to ensure data collection consistency while allowing for follow-up questions and probes to elicit more in-depth responses to the primary questions. We will consider emergent themes as well as unique comments, as some of our stakeholders may represent unique populations. We will consider stakeholder comments regarding different consumer populations (e.g., older adults, younger people with disabilities, etc.), different kinds of provider organizations (e.g., nursing homes, in-home care providers, medical day providers, etc.), and different kinds of community organizations (e.g., county welfare agencies, Area Agency on Aging, etc.) with respect to their ability to serve consumers. That is, we are interested in obtaining from our interviewees a picture of the processes through which consumers progress as they access Medicaid long-term services and supports—from information and referral, eligibility determination and redetermination (financial and clinical), MCO enrollment, care planning, receipt of services, handling of transitions due to clinical or social changes with regard to the consumer, and other issues that may be mentioned. We will identify themes and patterns in the interviews using an inductive process. Ongoing analysis of completed interviews will inform subsequent interviews with respect to follow-up questions.

Cost-Effectiveness

The evaluation will examine a robust set of measures of provider access and clinical quality to determine the cost-effectiveness of the demonstration policies. We will consider selected outcome measures included above relating to each evaluation hypothesis. We will utilize the results from regression analysis modeling the effect of the policy on such outcomes to assess the magnitude of changes in outcomes due to the policy change relative to a comparison population that was not subject to the policy.

Cost effectiveness methods will be based on best practices set forth by the 2nd US Panel in Cost Effectiveness in Health and Medicine (Neumann, 2016).[7] The primary cost-effectiveness measure for each intervention will be defined as the incremental cost effectiveness ratio (ICER), which represents the incremental difference between pre- versus post- policy costs divided by the difference in pre- versus post-policy outcome, for policies where a clear primary outcome can be defined.

$$ICER = \frac{\sum Cost_{post\text{-}policy} - \sum Cost_{pre\text{-}policy}}{\sum Outcome_{post\text{-}policy} - \sum Outcome_{pre\text{-}policy}}$$

[7] Neumann PJ, Sanders GD, Russell LB, Siegel JE, and Ganiats TG. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press, 2016. Second Edition

The numerators, $\sum Cost_{post\text{-}policy}$ and $\sum Cost_{pre\text{-}policy}$ represents the sum of total costs during the post-policy period, and total costs during the pre-policy period, respectively, and the denominator represents the sum of total outcome gained (or lost) during the pre- versus post-period. Each ICER thus indicates the additional costs to bring about one additional unit of benefit (outcome) from the policy. Cost effectiveness will be calculated from the state's perspective. This perspective captures the direct costs paid by government healthcare purchasers. These direct costs may include long term care, hospitalizations, emergency room and urgent care visits, outpatient care and tests, durable medical equipment, and medications. Due to the lack of data available on indirect costs such as productivity of the care recipient and productivity of the caregiver, it is not possible to conduct a societal cost effectiveness analysis.

Subject to availability of such information, costs of the policy change itself will be calculated using wage rates for personnel multiplied by time in preparation, documentation, training and supervision by adapting a model previously employed for CEA of a community-based intervention by the economic investigators.[8] Fringe benefit costs will be added to staff member costs by application of the prevailing state fringe benefit rate. Total costs of the policy intervention, reported in dollars during the year of implementation, will be defined as the sum of five direct cost categories; internal (e.g., staff) and external (e.g., organizations affected by and/or implementing the policy) training, intervention materials, staff travel associated with training and/or implementation of the policy change, and supervision/adherence of the policy change. The value of interventionist time will be calculated as the present value of earnings, and will be calculated as: (number of hours spent on the policy change task) x (interventionist's reported wage rates + fringe benefits). Staff training time for interventionists will be captured and converted to costs based on application of hourly wage rates as above. Material costs will include brochures, documentation forms and other education print and online materials provided to study participants. Staff travel expenses associated with the policy change will be costed based on reimbursement at the government rate (which will be obtained at time of the cost analysis but is expected to approximate $0.55 a mile).

The resulting ICERs we obtain will be examined relative to the previously reported willingness-to-pay thresholds *as available*. Willingness to pay thresholds using the standard metric (which is cost per quality-adjusted life year and ranges from $50,000-$100,000/quality adjusted life year in the US) will not be available since quality adjusted life years (QALYs) are not captured in the data and further, the methods of capturing QALYs in persons with disabilities may require proxy measurement from a caregiver who

[8] Gitlin LN, Harris LF, McCoy M, Chernett NL, Jutkowitz E, Pizzi LT. A community-integrated home based depression intervention for older African Americans: description of the Beat the Blues randomized trial and intervention costs. BMC Geriatr 2012;12:4.

may or may not have sufficient information and experience with the care recipient to accurately report quality adjusted life. Instead we anticipate the effectiveness measures in our cost effectiveness analyses to be clinical quality measures and/or care process measures. For example, a cost effectiveness analysis for diabetes could reasonably employ a measure of cost per individual achieving HbA1c value ≤ 7% since HbA1c targets are evidence-supported measures pertaining to diabetes control and risk of long-term complications. Our effectiveness measure will thus need to be tailored for each CEA and based on evidence-supported outcomes which are meaningful to the intervention being evaluated.

Sensitivity analyses will be conducted in order to determine the robustness of the ICERs. Both univariate sensitivity analysis (whereby one variable is changed at a time and impact on the ICER is examined), and probabilistic sensitivity analysis (PSA, whereby all relevant variables are simultaneously modified within reasonable ranges) will be conducted. Sensitivity analyses will include those variables where we anticipate “real world” uncertainty.

We will assess and compute all available costs associated with each policy change. When it is not possible to assess cost-effectiveness for lack of information on outcomes, we will assess whether there is any cost-savings as a result of the policy. Costs assessed over multiple periods will be inflation-adjusted (using the medical care price index) and subject to an appropriate discounting factor.

II. Evaluation of the New Jersey Delivery System Reform Incentive Payment (DSRIP) Program

BACKGROUND AND AIMS

The DSRIP is a component of the New Jersey Medicaid Comprehensive Waiver Demonstration initially implemented over the period October 2012 to July 2017. Under the Waiver renewal, the DSRIP program will continue for a period of three years over August 1, 2017 to June 30, 2020. The evaluation will examine the impact across all demonstration years, but distinguishing the effects by the first and the second round of the program, in accordance with the evaluation questions 10 and 11 that are stated in the special terms and conditions document. These are:

Was the DSRIP program effective in achieving the goals of better care for individuals (including access to care, quality of care, health outcomes), better health for the population, or lower cost through improvement? To what degree can improvements be attributed to the activities undertaken under DSRIP?

What do key stakeholders (covered individuals and families, advocacy groups, providers, health plans) perceive to be the strengths and weaknesses, successes and challenges of the expanded managed care program, and of the DSRIP pool? What changes would these stakeholders recommend to improve program operations and outcomes?

The evaluation questions for the DSRIP program based on the DSRIP planning protocol and the special terms and conditions documents relating to the first demonstration period, were the following:

1. To what extent does the program achieve better care?
2. To what extent does the program achieve better health?
3. To what extent does the program lower costs?
4. To what extent did the program affect hospital finances?
5. To what extent did stakeholders report improvement in consumer care and population health?
6. How do key stakeholders perceive the strengths and weaknesses of the program?

As we see above, the evaluation questions for the waiver renewal are identical to those for the first round of evaluation with the sole exception being one question related to the program impact on hospital finances. The stakeholder interviews in the first round also invited views and opinions on improving program implementation, an aspect that is explicitly mentioned in the current set of evaluation questions. Accordingly the evaluation methods for the DSRIP renewal will remain largely unchanged from those in

the first round, but there are three enhancements in the analytic strategy. First, we will take into account that comparison groups may be systematically different from DSRIP adopting hospitals and conduct additional analysis to account for these differences. Second, as mentioned above, we will model differences in program impact between the first and second rounds of demonstration. Finally, in addition to the Medicaid fee-for-service and managed care encounter data that we receive from the state, we will additionally use all-payer hospital discharge data to examine DSRIP effects among the uninsured population. Greater details regarding all of these plans and associated identification strategies are provided in the analytic section below.

We begin by providing a brief background, followed by specific hypotheses related to the evaluation questions, description of data sources, outcomes, and statistical and econometrics techniques to identify program effects.

The DSRIP program uses resources from the previously existing hospital relief subsidy fund to establish a system of incentive payments for hospitals based on achieving specific health improvement goals. The stated goals of the program include “better care for individuals (including access to care, quality of care, health outcomes), better health for populations and lower cost through improvement.” In this population health management program, hospitals select specific disease management projects based on the needs of the populations served and are assessed on the basis of quality metrics that measure the effectiveness of their programs in improving access and quality of care and health outcomes.

The evaluation will examine the effectiveness of the DSRIP program overall and specific disease management programs. We formulated specific testable hypotheses related to DSRIP hospital programs, patient access and quality of care, patient health, costs of care, and stakeholder perceptions relating to the program that would answer these questions and ultimately shed light on the effectiveness of the DSRIP program.

The five hypotheses along with their corresponding sub-hypotheses are detailed below. Appendix A1 presents a crosswalk between each of these hypotheses and the DSRIP research question(s) (enumerated above) that it addresses. Below each hypothesis we categorize the measures that will be used to test it. Each category of measures represents one or more metrics that are detailed in Appendix A2 and Tables 1 and 2.

Hypothesis 1: The adoption of hospital projects in a specific focus area (e.g., cardiac care, asthma) will result in greater improvements in related care and outcomes for patients from hospitals adopting these interventions compared to hospitals which do not adopt these interventions.

This general hypothesis can be broken down into seven sub-hypotheses that examine the effectiveness of each of the seven chronic condition projects that include asthma;

behavioral health; cardiac care; chemical addiction/substance abuse; diabetes; obesity; and pneumonia. For instance,

Hypothesis 1a: Rates of 30-day heart failure/acute myocardial infarction readmissions will decrease in hospitals adopting cardiac care interventions during the DSRIP program.

Hypothesis 1b: Rates of asthma admissions and ED visits will decrease for patients in hospitals adopting asthma management programs.

Hypothesis 1c: Rates of follow-up visits after hospitalizations for mental illness will increase for patients from hospitals adopting behavioral health interventions during the DSRIP program.

Hypothesis 1d: Rates of initiation and engagement in alcohol and other drug treatment will increase for patients from hospitals adopting chemical addiction/substance use management projects during the DSRIP program.

Hypothesis 1e: Rates of admissions for diabetes short-term complications will decrease for patients from hospitals adopting diabetes management projects during the DSRIP program.

Hypothesis 1f: Rates of 30-day pneumonia readmissions will decrease for patients from hospitals adopting pneumonia intervention projects during the DSRIP program.

Hypothesis 1g: Rates of children's and adolescents' access to primary care practitioners will increase for patients from hospitals adopting obesity intervention projects under the DSRIP program.

As Appendix A1 outlines, hypothesis 1 addresses the research questions on whether the program achieves better care and outcomes by examining metrics relating to hospital admissions, readmissions, treat-and-release emergency department visits, and recommended care. (The specific metrics are detailed in the 'outcome variables' section in Methods, and also in Appendix A2 that relates each hypothesis to the specific metrics). The focus of hypothesis 1 is the effectiveness of the chronic disease management projects in the DSRIP program.

Hypothesis 2: The DSRIP program will improve the quality of ambulatory care in the communities of participating hospitals consequently reducing avoidable inpatient hospitalizations and avoidable/preventable emergency department visits; it will improve access to care; quality and efficiency of care.

Hypothesis 2 thus examines all three research questions relating to better care, better health and lower costs. The quality and adequacy of ambulatory care will be measured by avoidable inpatient hospitalizations and ED visits. These, and other hospital specific outcomes, and additional measures related to recommended care examine the impact

of the program on better care and better health in the population. Finally, a decrease in costs associated with avoidable hospitalizations would indicate increasing efficiencies in care.

Hypothesis 3: The DSRIP program will reduce racial/ethnic and gender disparities in avoidable hospital admissions, treat-and-release ED visits, and hospital readmissions, in participating hospitals.

Hypothesis 3 also sheds light on whether the program improves care and ensures better health in the population. This specifically recognizes the importance of ensuring that program benefits reach all sections of the Medicaid population. Hospitalizations stratified by race/ethnicity and gender will reveal whether readmission rates or ambulatory care sensitive hospitalizations are higher among racial/ethnic minorities and/or women.

Hypothesis 4: Stakeholders will report improvements in consumer care.

Hypothesis 5: Stakeholders will report improvements in population health.

Hypotheses 4 and 5 are tested through key informant interviews and examine whether stakeholders perceive that the DSRIP program will improve consumer care and population health. In order to shed light on such pathways, questions included in the interviews and surveys will also identify implementation experiences, positive or negative, that arise from program characteristics.

EVALUATION STRUCTURE AND PLANNING

Guided by the research questions and the corresponding hypotheses, the evaluation will examine the impact of the DSRIP program on patient care, patient health, and costs of providing care; it will also examine stakeholder perceptions relating to population health and overall strengths and weaknesses of the program. This evaluation will thus utilize a mix of quantitative and qualitative methods.

The quantitative component will provide an independent analysis of key metrics to inform how well the DSRIP Program achieves better care and better health for populations served by hospitals, as well as lower costs through improvement. Qualitative analysis, including key informant interviews and document review, will be conducted throughout planning and implementation of the DSRIP Program, to provide stakeholder perceptions of improvements in care and strengths and weaknesses of the program.

Quantitative process and outcome measures along with inputs from qualitative analyses will be utilized to independently analyze and interpret data evaluating hypotheses 1-3. A qualitative approach will answer questions 4 and 5 based on stakeholder interviews, observations of program meetings, and review of relevant documents.

The evaluation report will meet all standards of leading academic institutions and academic peer review, as appropriate for both aspects of the DSRIP program evaluation, including standards for the evaluation design, conduct, interpretation, and reporting of findings.

The single evaluation report examining the DSRIP program over January 1, 2014 to June 30, 2020 will be completed by the end of December 2021.[9]

QUANTITATIVE EVALUATION

APPROACH AND METHODS

Overall strategy and design

We will identify the effect of the DSRIP program on provision of care and population health by examining changes in specific healthcare and health related outcomes over time. These outcomes calculated through metrics detailed in Tables 1 and 2 will be based on Medicaid fee-for-service claims and managed care encounter data. We will also calculate select metrics based on all-payer hospital discharge data for the uninsured population.

We will use a difference-in-differences analysis for specifications where we can define a comparison group. Here, hospitals will be classified into study or comparison groups based on their participation in the DSRIP program and also individual disease-specific projects, each classification thus varying, depending on the category of the hypothesis being tested (effectiveness of individual programs or success of the overall DSRIP program) The differences in trends (in hospital performance captured through the metrics) between the study and comparison group from the baseline (2011-2013) to the first implementation period (2014-2017) to the second implementation period (2017-2020) will identify the program effects.

We will also utilize interrupted time series modeling that does not require a comparison group.

See details regarding how these methods will be implemented in the analytic section below.

Data:

Sources: The evaluation team will independently calculate evaluation-related measures using NJ Medicaid fee-for-service claims along with managed care encounter data. We will additionally use all-payer hospital discharge data to examine program effects on the uninsured population.

[9] This timeline is contingent on timely receipt of Medicaid claims/encounter data from DHS.

Availability: Medicaid-paid fee-for-service claims and encounter data will be available from Medicaid during the period of the evaluation. Monthly extracts are received and used to build static analytic claims files. The State has estimated that the majority of FFS and managed care claims are received within six months of the date of service, and we will apply a Medicaid-recommended lag period of at least six months to allow for retroactive adjustments to the data. This will allow accurate measurement of costs and payments and also provide consistency and comparability with other parts of the evaluation. Our analytic files are validated against a real-time database query from DMAHS on total payment amounts, total number of claims, and recipient eligibility counts for a specified period and differ by <1%. Due to this adjustment period and also the time required to analyze data and statistically model evaluation effects, there will be a period of delay from the end of the DSRIP demonstration until the availability of the evaluation report.

All-payer hospital discharge data is available from AHRQ HCUP state inpatient databases (SID) and state emergency department databases (SEDD). If HCUP data are used, the latest year available for our evaluation report will be 2018. We are in discussion with the state of New Jersey on the availability of linked discharge data that will also allow us to calculate metrics that require patients to be followed over time (e.g., readmissions) in addition to point-in-time metrics (e.g., avoidable inpatient stays and ED visits). If data are received directly from the State, data through 2019 may be available.

Outcome variables

The metrics related to our outcomes of interest are detailed in Tables 1 and 2. The first category of metrics included in Table 1 examines effectiveness of hospital-specific chronic condition projects and allows testing of hypothesis 1 and its seven sub-hypotheses. For instance, an increase in follow-up visits after hospitalizations for mental health indicates the effectiveness of behavioral health programs being pursued by some hospitals. The second category of outcomes/metrics listed in Table 2 test the remaining hypotheses assessing the overall impact of the DSRIP program - on quality and efficiency of care within the delivery system, patient health, and racial and ethnic disparities in care. For instance, did avoidable hospitalizations and ED visits that arise from inadequate ambulatory care in the community decrease; did rates of 30-day all-cause readmissions among patients admitted for heart attack, heart failure or pneumonia decrease among DSRIP hospitals?

Appendix A2 gives detailed definitions for calculating these metrics which are of two types, hospital-event based metrics and population-based metrics. The former, such as hospital readmission rates, will be calculated at the hospital level based on all discharges from specific hospitals. For population-based metrics (e.g., rates of avoidable inpatient hospitalizations, ED visits rates for asthma, and rates of patients receiving substance use related treatment), we will calculate zip code population-based rates and then classify those zip codes based on whether the hospitals serving the majority of patients residing there took part in specific DSRIP programs.

Appendix A2 also links each of these metrics to measure domains that enables testing one or more of the three hypotheses related to the quantitative evaluation. The domains are outcomes from the chronic disease programs (Hypothesis 1); additional health outcomes (Hypothesis 2); care processes that capture access to quality care and preventive/recommended care (Hypothesis 2); and racial/ethnic disparities (Hypothesis 3). Some of the metrics may address multiple hypotheses. Diabetes short-term complication admission rate examines the effectiveness of hospital diabetes programs (Hypothesis 1). In addition, being an ambulatory care sensitive condition, it sheds light on improvements in access and quality of care in the community (Hypothesis 2).

While selecting our metrics we chose such measures that reflect the effect of the intervention on the overall delivery system, those that assess inpatient as well as ambulatory care received by patients, in contrast to much narrower inpatient process measures which are further removed from patient outcomes. Metrics were also specifically chosen to reflect the current policy changes related to hospital financing, such as rates of all-cause readmissions from initial hospitalizations of heart failure, AMI and pneumonia. We adopted definitions posted by organizations such as NQF and NCQA; however, it may be necessary to adapt some of those criteria to the evaluation objectives and data availability. An underlying criterion during the metric selection process was to choose measures that can be independently calculated by the evaluator from claims/encounter-based data. Metrics that require medical charts and cannot be independently calculated (e.g., those related to screening for depression) do not fall in this category.

Table 1: Metrics for evaluating hospital specific projects

	Metric
Asthma	Percent of patients who have had a visit to an Emergency Department (ED) for asthma in the past six months.[a] *Adult Asthma Admission Rate**
Behavioral Health	Follow-up After Hospitalization for Mental Illness (30 days post discharge) *Follow-up After Hospitalization for Mental Illness (7 days post discharge)*
Cardiac Care	30-Day All-Cause Readmission Following Heart Failure (HF) Hospitalization *30-Day All-Cause Readmission Following Acute Myocardial Infarction (AMI) Hospitalization*
Chemical Addiction/ Substance Abuse	Engagement of alcohol and other drug treatment *Initiation of alcohol and other drug treatment*
Diabetes	Diabetes Short-Term Complications Admission Rate* *Comprehensive Diabetes Care: Hemoglobin A1C testing* Comprehensive Diabetes Care: Eye exam (retinal) performed
Pneumonia	30-Day All-Cause Readmission Following Pneumonia (PN) Hospitalization
Obesity	Children and Adolescents' Access to Primary Care Practitioners

All metrics will be calculated using FFS claims and managed care encounter data.

*Metric will also be calculated in all-payer hospital discharge data for the uninsured population.
[a]original metric included visits to urgent care office; which cannot be identified in Medicaid claims/encounter data.

Table 2: Metrics for Overall Evaluation of the DSRIP Program

	Description
Mental Health Utilization	The number and percentage of patients receiving inpatient mental health services during the measurement year.
30-Day All-Cause Readmission Following Heart Failure (HF) Hospitalization	The measure estimates a hospital-level, risk-standardized, all-cause 30-day readmission rate for patients discharged from the hospital with a principal discharge diagnosis of Heart Failure (HF).
30-Day All-Cause Readmission Following Acute Myocardial Infarction (AMI) Hospitalization	The percent of 30 day all-cause readmission rate for patients with AMI.
30-Day All-Cause Readmission Following Pneumonia (PN) Hospitalization	The percent of 30 day all-cause readmission rate for patients with pneumonia.
30-Day All-Cause Readmission Following Chronic Obstructive Pulmonary Disease (COPD) Hospitalization	The percent of 30 day all-cause readmission rate for patients with COPD.
Rate of potentially avoidable inpatient hospitalizations reflecting inadequate level of ambulatory care. Based on AHRQ methodology for calculating Prevention Quality Indicators.*,[10]	
Rate of Primary Care Preventable/Avoidable Treat and Release ED visits. Based on methodology by John Billings, New York University.*,[11]	
Hospital costs related to avoidable inpatient stays, and treat-and-release Emergency Department visits	
Well Child Visits in the First 15 Months of Life	Percentage of patients who turned 15 months old during the measurement year and who had well-child visits with a PCP during their first 15 months of life
Emergency Department Visits*	Rates of treat-and-release emergency department visits

All metrics will be calculated using FFS claims and managed care encounter data.
*Metric will also be calculated in all-payer hospital discharge data for the uninsured population.

[10] Bindman AB, K Grumbach, D Osmond, M Komaromy, K Vranizan, N Lurie, J Billings, and A Stewart. "Preventable Hospitalizations and Access to Health Care." *Journal of the American Medical Association* 274, no. 4 (1995): 305–11.

[11] Billings J, N Parikh, and T Mijanovich. *Emergency Department Use: The New York Story*. New York: Commonwealth Fund, 2000.

Analytic Strategies to Identify Policy Effect

Difference-in-Differences Approach: The evaluation will utilize a difference-in-differences (DD) estimation technique that examines changes in the levels and trends of selected outcomes before and after the implementation of the program/policy comparing DSRIP hospitals in specific programs and comparison hospitals. Such an estimation strategy is able to identify the changes in outcomes that are due to program impact, and distinct from secular trends in outcomes that are unrelated to our policy of interest.

The DD strategy examines the effectiveness of the individual chronic disease management programs as well as the DSRIP program overall in improving care and health by comparing specific metrics (from Tables 1 and 2) for study and comparison hospitals over time. For the first hypothesis, the study group comprises hospitals taking part in specific projects (cardiac care) and comparison group comprises hospitals not taking part in those projects. Project-specific outcomes (e.g., rates of heart failure readmissions) are compared between patients in the study hospitals to those in comparison hospitals in the pre- and post-policy periods. In order to implement this approach, the selected project-specific metrics (see Table 1) will be calculated for all hospitals. For example, rates of heart failure admissions will be calculated for all hospitals, comparing hospitals that selected cardiac care as their DSRIP focus (study group) to those which did not (comparison group). For the remaining hypotheses examining the overall impact of the DSRIP program, all hospitals approved for the DSRIP program will constitute the study group and will be compared to all remaining acute-care hospitals in New Jersey. Over the course of the program, the number of hospitals in the comparison group may increase if some hospitals decide to discontinue participation in the program. Our data analysis will incorporate such changes.

$$Y_{it} = \beta_0 + \beta_1(program)_i + \beta_2(post\_1)_t + \beta_3(post\_2)_t + \beta_4(program_i * post\_1_t) + \beta_5(program_i * post\_2_t) + \gamma X_{it} + \varepsilon_{it} \quad (1)$$

The variable Y_{it} represents the outcome for the ith hospital or zip code depending on the specific outcome, at year t. Post_1= 0 or 1 depending on whether the time is during the first round of the DSRIP program (January 1, 2014- July 31, 2017), post_2=0 or 1 depending on whether the time is during the second round of the demonstration (August 1, 2017- June 30, 2020). The reference category is the baseline period spanning January 1, 2011- December 31, 2013. The statistical model in equation (1) thus accounts for these three distinct periods by incorporating the indicator variables for specific years or rounds of demonstration. This will allow estimation of changes in outcomes during the first DSRIP demonstration period from the policy implementation, and additional changes in outcomes during the second demonstration period from continuation of those policy changes. In the case of a hospital based metric, program =1, if the hospital is taking part in the DSRIP program, 0 otherwise. In case of an outcome metric that has a population-based denominator, the unit of analysis is a zip

code and we will follow methods[12] previously developed at Rutgers CSHP. Here, for our baseline specification, program=1 if at least one of the hospitals serving the patients residing in that zip code are taking part in the program; in alternative specifications, program will be a continuous variable reflecting the share of patients belonging to DSRIP hospitals out of the "relevant" set of hospitals serving a zip code. This relevant set of hospitals will comprise the smallest set that account for 75% or more of the total inpatient and ED volume from that zip code. Additional sensitivity analysis will define the relevant set of hospitals based on thresholds of 50% and 90% of total volume of patients from zip codes. We will adopt identical strategies while modeling the effect of a specific DSRIP program.

X is a vector of other control variables relating to patient, zip code and hospital level characteristics. Depending on whether the outcome is assessed at the zip code or hospital-level, we will include zip code or hospital fixed effects[13]. ε_{it} represents the random error term.

In this specification β_5 measures the program impact during the second round of demonstration relative to the baseline period and β_4 measures program impact during the first round of the demonstration, also relative to the baseline period. The difference between these effect sizes will provide the incremental impact of the policy during the second round relative to the first round.

Depending on the specific measure, Y_{it} can be a rate or a binary or count variable, and appropriate functional forms (e.g., ordinary least square, logistic, linear probability model, Poisson, negative binomial) will be chosen accordingly. For example, a logistic specification utilizing a discharge-level analysis may be used to estimate the effect of the program on the likelihood of a patient being readmitted within 30 days. In case of a population-based measure such as asthma admissions, the analysis will be at the zip code level. The outcome variable would be total asthma admissions from patients in a zip code per zip code population. The zip code will be classified based on whether the hospitals serving that zip code took part in asthma management project. Spending will be modeled using a gamma distribution with a log link specification.

The overarching goal of these methods is to support measurement of the impact of these programs on the demonstration goals, examine causal pathways by identifying confounders and accounting for the effect of other interventions in the state that may have interacted with this demonstration, such as the implementation of the Accountable Care Organizations and the effect of 2014 Medicaid expansion.

[12] DeLia, D., Cantor, J. C., Tiedemann, A., & Huang, C. S. (2009). Effects of regulation and competition on health care disparities: the case of cardiac angiography in New Jersey. *J Health Polit Policy Law, 34*(1), 63-91.

[13] See details regarding these methods in our midpoint and final evaluation of the NJ DSRIP program.

Examining suitability of comparison groups: DD modeling identifies the impact of the policy change by comparing the trend in outcomes for the study population from the pre- to the post-implementation period(s) to that of a comparison group which is otherwise similar, but not subject to the policy effect. The DD estimate is able to account for the effect of unobserved factors and generate an estimate of the true policy effect as long as the impact of the policy on the intervention group relative to the comparison group does not change over time. We will test this by examining whether trends in outcomes prior to policy implementation (pre-trends) for the intervention and comparison group are parallel to each other. Each regression model will examine in supplementary analysis whether there exist statistically significant differences in trends between the intervention and comparison group prior to policy implementation. If this difference is in the same direction as the DD estimate and of comparable magnitude that would imply that the DD model may be overestimating the effect. Accordingly our estimation process of computing effect sizes will adjust for these differential effects based on well-established methods in peer-reviewed academic publications.[14]

Potential differences between intervention and comparison groups: There may be systematic differences between hospitals taking part in certain projects and those that are not. Further such differences may also exist between the communities served by these hospitals. This is because hospitals may choose to implement projects that are relevant to the patients that they serve and/or where they have prior experience and expertise. In our descriptive analysis, we will examine and report outcomes as well as differences in provider and patient characteristics between treatment and comparison hospitals to see whether they are significantly different. It is important to note that DD estimates are valid even when outcomes for program hospitals (even before policy implementation) are systematically different from those of comparison hospitals (which may be the case because of reasons described above) as long as the trends in outcomes are parallel to each other. As mentioned above, we will examine and account for such differences in pre-trends based on academic publications and our previous work.[15,16]

[14] Harman, J. S., Hall, A. G., Lemak, C. H., & Duncan, R. P. (2014). Do provider service networks result in lower expenditures compared with HMOs or primary care case management in Florida's Medicaid program? Health Serv Res, 49(3), 858-877. PMCID: PMC4231575

[15] Akosa Antwi, Y., Moriya, A. S., Simon, K., & Sommers, B.D. (2015). Changes in Emergency Department Use Among Young Adults After the Patient Protection and Affordable Care Act's Dependent Coverage Provision. Ann Emerg Med, 65(6), 664-672. PMCID: PMC 2576946

[16] Chakravarty, S., Lloyd, K., Farnham J., Brownlee, S., & DeLia D. (2017). Examining the Effect of the NJ Comprehensive Waiver on Access to Care, Quality, and Cost of Care: Draft Final Evaluation Report. New Brunswick, New Jersey: Rutgers Center for State Health Policy. Available at:

Interrupted time series modelling: While we will develop comparison groups wherever feasible in our evaluation analyses to facilitate separation of program impact from secular trends in outcomes, it may not be always possible to have suitable comparison groups. This may be because of systematic differences between intervention and comparison groups discussed above or due to inadequate sample size of non-participating hospitals. For those measures, segmented regression analysis/interrupted time series modeling will be used to allow inferences about DSRIP impact. Such a model assumes that the policy effect may lead to a change in level, and also a change in the existing time trend of the metric measuring quality or any other relevant outcome of interest. The regression analysis is able to measure this change in trend or level. Potential confounding may arise in the rare circumstances when policy-unrelated factors that determine our outcomes of interest change at exactly the same time as the policy implementation. However, our multivariate analysis adjusting for patient, provider and geographic factors are expected to mitigate such effects. The model also allows us to account for policy changes occurring in multiple points of time. Equation (2) below represents such a model based on our previous work.[17]

$$Y_{it} = \beta_0 + \beta_1(time)_t + \beta_2(DSRIP\_1\ post)_t + \beta_3(DSRIP\_1\ time)_t + \beta_4(DSRIP\_2\ post)_t) + \\ +\beta_5(DSRIP\_2\ time)_t + \gamma X_{\ it} + \varepsilon_{it} \qquad (2)$$

Here, Y_{it} reflects the outcome related to the i^{th} hospital or zip code at time *t*. On the right hand side of the equation, time is a continuous variable indicating time in months or calendar quarters from the start of the study period i.e., January 2011. The variables *dsrip_1 post* and *dsrip_2 post* are indicator (0/1) variables for the period during the first and second round of DSRIP implementation. The variables *dsrip_1 time* and *dsrip_2 time* are continuous variables equaling the number of months (or quarters) after the start of the first and second rounds of DSRIP implementation. Patient, provider and zip code characteristics are represented by the variable X_{it}. ε_{it} is the random error term utilized in the regression representing the statistical distribution of the outcome variable.

Coefficient β_0 estimates the baseline level of the outcome coefficient β_1 indicates the baseline trend prior to the first round of DSRIP. Coefficients β_2 and β_4 estimate the level changes after the initiation of each round of DSRIP in January 2014 and July 2017

http://www.cshp.rutgers.edu/publications/examining-the-effect-of-the-nj-comprehensive-waiver-on-access-to-care-quality-and-cost-of-care-draft-final-evaluation-report.

[17] Chakravarty, S., Lloyd, K., Farnham J., Brownlee, S., & DeLia D. (2017). Examining the Effect of the NJ Comprehensive Waiver on Access to Care, Quality, and Cost of Care: Draft Final Evaluation Report. New Brunswick, New Jersey: Rutgers Center for State Health Policy. Available at: http://www.cshp.rutgers.edu/publications/examining-the-effect-of-the-nj-comprehensive-waiver-on-access-to-care-quality-and-cost-of-care-draft-final-evaluation-report.

respectively. Similarly β_3 and β_5 estimate the change in trend in the outcome after each of these policy changes. The specification detailed above, is able to identify changes in outcomes that may have occurred due to the first round of DSRIP implementation and isolate those effects from that of second round of DSRIP implementation.

As an illustrative example, the specific effect of the second round of DSRIP is given by the magnitude of β_4 that gives the change in level and β_5 that gives the change in trend after the DSRIP implementation and we further test whether these values are statistically significant. Accordingly in our results section, we will report the magnitudes of these two coefficients and their joint statistical significance. For interpretability purposes, we will further compare predicted values of outcomes post-DSRIP with counterfactual values (that simulate a scenario where the DSRIP implementation did not occur). We will further compute whether this difference is statistically significant.

Adjusting for Patient, Provider and Geographic Factors: As demonstrated in the different model specifications, our analysis will control for patient characteristics that may affect outcomes. These include beneficiary demographics, Medicaid eligibility category, health history (including chronic illness and behavioral health co-morbidities), chronic disability payment score, and any other information relevant to the policy of interest. We will incorporate hospital fixed effects (to account for time-invariant differences across hospitals) for inpatient quality-based measures and zip code fixed effects (to account for time-invariant measures across geographic locations) for measures reflecting ambulatory care.

For specific outcomes that reflect the overall delivery system (e.g., avoidable hospitalizations and readmissions) analysis will examine differences across patient populations differentiated by race/ethnicity and gender to the extent that sample sizes permit. Because of the diversity of the New Jersey population, we expect to find differences in the effect of the DSRIP program among demographic groups and we will document these differences.

Sensitivity Analysis: We will also conduct sensitivity analysis through a falsification test that estimates a placebo model by falsely assuming that the policy change was implemented in 2013. Based on methods previously used by evaluation researchers[18], this examines whether there were any statistically significant changes in outcomes, one year prior to the DSRIP implementation.

[18] Cantor, J.C., Monheit, A.C., DeLia, D. and Lloyd, K. (2012). Early impact of the affordable care act on health insurance coverage of young adults. Health Serv Res, 47(5), 1773-90.

We will add a test examining outcomes not expected to be affected by the DSRIP program. Some candidate outcome measures would be annual dental visits, substance-use related hospitalizations (for hospitals not conducting chemical addiction/substance use projects), and hospitalizations for epilepsy.

Our estimation procedures will be conducted using standard inferential statistical techniques employing STATA 15.0 or SAS 9.2 software.

QUALITATIVE EVALUATION

This section below describes the qualitative methods used to gather and analyze data to examine stakeholder perceptions relating to the DSRIP program and address hypotheses 5 and 6.

To address research questions 5 and 6 and test hypotheses 4 and 5, related to stakeholder perceptions, the evaluation team will develop an interview protocol to gather views of stakeholder perceptions about DSRIP program effectiveness in improving access, quality of care, and population health outcomes. The interviews will take place over January-June 2020. We conduct this during the last six months of the program anticipating personnel changes once the program ends and difficulty in identifying interviewees.

To provide background for the stakeholder-directed questions, the evaluation team will also review information available from hospital projects, such as program materials, community outreach materials, presentations, and reports from participating hospitals. The interview protocol will be approved by the Rutgers University Institutional Review Board, and interviewers will be trained to ensure privacy and confidentiality.

The evaluation team will gather information regarding the questions detailed below, as well as others suggested by DSRIP stakeholders.

- What positive impacts did you observe from the DSRIP project? Which patient and/or community groups experienced benefits? Were these the expected groups?
- What difficulties were encountered in developing and sustaining a DSRIP project, e.g., obtaining resources, engaging community partners, collecting and sharing clinical data, etc.? How were difficulties addressed? Which strategies were most successful? What additional information would have been helpful in carrying out the DSRIP program?
- What difficulties were encountered in implementation of the DSRIP project?
- What changes in policy or practice external to the DSRIP have affected implementation of the DSRIP or made it difficult to gather accurate information?
- What problems or improvements in consumer care have been noted in your community?
- What problems or improvements in the health of specific population groups have been noted in your community?
- What improvements in health care were made as a result of the DSRIP projects?

- What new clinical partnerships were developed?
- How were real time data used to support the efforts of hospitals to refine their programs?
- How did the learning collaborative support change? What could have made the Learning Collaborative more successful?
- What other rapid-cycle improvement tools were used and how effective were they in supporting quality improvement? Was there adequate support for hospitals for these activities? What could make the rapid-cycle tools (e.g. learning collaborative, dashboards, real time data exchanges, etc.) more effective?
- Were there unanticipated consequences in hospital operations, other programs, or financial status?

Key informant interviews will be conducted with officials from the Department of Health and the Department of Human Services, as well as other stakeholders familiar with the program including representatives from hospital associations. Interviews will also be conducted with representatives from hospitals' community partners to obtain viewpoints about expected benefits and unanticipated consequences for patients and families.

Interviewers will use a semi-structured guide containing key questions to ensure data collection consistency while allowing for follow-up questions and probes to elicit more in-depth responses to the primary questions. Data from key informant interviews will be transcribed and de-identified, then independently coded by two researchers to identify themes and patterns in the data. We will specifically compare safety-net and non safety-net hospitals and consider interviewee comments regarding differential effects of the program on different communities or groups of patients. Ongoing analysis of completed interviews will inform subsequent interviews.

Appendix A1: Crosswalk Between Research Questions and Proposed Evaluation Hypotheses

Evaluation Hypotheses & Measure Domains[1]	Planning Protocol Research Questions[2]
Hypothesis 1: Hospital Projects improve related care and outcomes - hospital admissions (2,9) - hospital readmissions (5,6,10) - ED visits (1) - recommended care (3,4,7,8,11,18,19)	1. To what extent does the program achieve better care? 2. To what extent does the program achieve better health?
Hypothesis 2: Program improves quality of ambulatory care; recommended and preventive with positive effects on population health - avoidable inpatient hospitalizations (14) - avoidable/preventable ED visits (15) - ED visits (20) - associated costs (17) - recommended care (11,12,16,18,19) - hospital readmissions (5,6,10,13)	1. To what extent does the program achieve better care? 2. To what extent does the program achieve better health? 3. To what extent does the program lower costs?
Hypothesis 3: The DSRIP program will reduce racial/ethnic and gender disparities in avoidable hospital admissions, treatand release ED visits, and hospital readmissions. - avoidable hospitalizations stratified by race/ethnicity and gender (14,15) - hospital readmission rates stratified by race/ethnicity and gender (5,6,10,13)	1. To what extent does the program achieve better care? 2. To what extent does the program achieve better health?
Hypothesis 4: Stakeholders will report improvements in consumer care - perceived improvements in consumer care - implementation difficulties that may modify program impact	5. To what extent did stakeholders report improvement in consumer care and population health? 6. How do key stakeholders perceive the strengths and weaknesses of the program?
Hypothesis 5: Stakeholders will report improvements in population health - benefits experienced by patient or community groups - implementation difficulties that may modify program impact - new clinical partnerships with beneficial impact on population health	5. To what extent did stakeholders report improvement in consumer care and population health? 6. How do key stakeholders perceive the strengths and weaknesses of the program?

[1]Numbers in parentheses after the measure domain refer to the specific metric numbers as detailed in Appendix A2.

Appendix A2: Crosswalk Between Metrics and Evaluation Hypotheses

Metric Number	Evaluation[1]	Source	Metric Name	Metric Description	Hypothesis			
					1	2		3
					Chronic Disease Outcomes	Health Outcomes	Care	Disparities
1	ASTHMA		Percent of patients who have had a visit to an Emergency Department (ED)/Urgent Care office for asthma in the past six months.	This measure is used to assess the percent of patients who have had a visit to an Emergency Department (ED)/Urgent Care office for asthma in the past six months.	X			
2	ASTHMA	Medicaid Adult Core #11; PQI 15; NQF 0283	Adult Asthma Admission Rate (PQI-15)	This measure is used to assess the number of admissions for asthma in adults under the age of 40 per 100,000 population.	X	X	X	
3	BEHAVIORAL HEALTH	HEDIS; Medicaid Adult Core #13; Medicaid Child Core; NQF 0576	Follow-up After Hospitalization for Mental Illness 30 days post discharge	The percentage of discharges for members 6 years of age and older who were hospitalized for treatment of selected mental health disorders and who had a follow-up visit with a mental health practitioner within 30 days of discharge.	X		X	
4	BEHAVIORAL HEALTH	HEDIS; Medicaid Adult Core #13; Medicaid Child Core; NQF 0576	Follow-up After Hospitalization for Mental Illness 7 days post discharge	The percentage of discharges for members 6 years of age and older who were hospitalized for treatment of selected mental health disorders and who had a follow-up visit with a mental health practitioner within 7 days of discharge.	X		X	
5	OVERALL & CARDIAC CARE	Joint Commission National Hospital Inpatient Quality Measures; NQF 0330	30-Day All-Cause Risk-Standardized Readmission Rate Following Heart Failure (HF) Hospitalization	The measure estimates a hospital-level, risk-standardized, all-cause unplanned 30-day readmission rate for patients discharged from the hospital with a principal discharge diagnosis of Heart Failure (HF).	X	X		X
6	OVERALL & CARDIAC CARE	Joint Commission National Hospital Inpatient Quality Measures; NQF 0505	30-Day All-Cause Risk-Standardized Readmission Rate Following Acute Myocardial Infarction (AMI) Hospitalization	The measure estimates a hospital-level, risk-standardized, all-cause unplanned 30-day readmission rate for patients discharged from the hospital with a principal discharge diagnosis of Acute Myocardial Infarction (AMI).	X	X		X

Metric Number	Evaluation[1]	Source	Metric Name	Metric Description	Chronic Disease Outcomes	Health Outcomes	Care	Disparities
					Hypothesis			
					1	2		3
7	CHEMICAL ADDICTION/ SUBSTANCE ABUSE	HEDIS; Medicaid Adult Core #25; NQF 0004	Initiation of alcohol and other drug treatment	This measure is used to assess the percentage of adolescent and adult members with a new episode of alcohol or other drug (AOD) dependence who initiate treatment through an inpatient AOD admission, outpatient visit, intensive outpatient encounter, or partial hospitalization within 14 days of the diagnosis.	X		X	
8	CHEMICAL ADDICTION/ SUBSTANCE ABUSE	HEDIS; Medicaid Adult Core #25; NQF 0004	Engagement of alcohol and other drug treatment	This measure is used to assess the percentage of adolescent and adult members with a new episode of alcohol or other drug (AOD) dependence who initiated AOD treatment and who had two or more inpatient admissions, outpatient visits, intensive outpatient encounters, or partial hospitalizations with any AOD diagnosis within 30 days after the date of the Initiation encounter (inclusive).	X		X	
9	DIABETES	Medicaid Adult Core #8; PQI 01; NQF 0272	Diabetes Short-Term Complications Admission Rate (PQI-01)	The number of discharges for diabetes short-term complications per 100,000 age 18 years and older population in a Metro Area or county in a one year period.	X	X	X	
10	OVERALL & PNEUMONIA	Joint Commission National Hospital Inpatient Quality Measures; NQF 0506	30-Day All-Cause Risk-Standardized Readmission Rate Following Pneumonia (PN) Hospitalization	The measure estimates a hospital-level, risk-standardized, all-cause unplanned 30-day readmission rate for patients discharged from the hospital with a principal discharge diagnosis of Pneumonia (PN).	X	X		X
11	OVERALL & OBESITY	HEDIS; Medicaid Child Core	Children and Adolescents' Access to Primary Care Practitioners	The percentage of patients 12 months–19 years of age who had a visit with a PCP. -Children 12–24 months and 25 months–6 years who had a visit with a PCP during the measurement year -Children 7–11 years and adolescents 12–19 years who had a visit with a PCP during the measurement year or the year prior to the measurement year	X		X	
12	OVERALL	HEDIS	Mental Health Utilization - Inpatient	The number and percentage of members receiving inpatient mental health services during the measurement year.			X	

Metric Number	Evaluation[1]	Source	Metric Name	Metric Description	Chronic Disease Outcomes	Health Outcomes	Care	Disparities
					Hypothesis			
					1	2		3
13	OVERALL	NQF 1891	30-Day All-Cause Risk-Standardized Readmission Rate Following Chronic Obstructive Pulmonary Disease (COPD) Hospitalization	The measure estimates a hospital-level, risk-standardized, all-cause unplanned 30-day readmission rate for patients discharged from the hospital with a principal discharge diagnosis of Chronic Obstructive Pulmonary Disease (COPD).		X		X
14	OVERALL	PQI 90	Preventable Hospitalizations	AHRQ created Prevention Quality Indicators (PQI) that are rates of potentially avoidable hospitalizations for ambulatory care sensitive conditions that reflect issues of access to, and quality of, ambulatory care in a given geographic area.		X	X	X
15	OVERALL		Preventable/Avoidable Treat and Release ED Visits	Based on methodology of John Billings at New York University, determines the proportion of treat-and-release ED visits that are: -Non-emergent -Emergent/primary care treatable -Emergent - ED Care Needed - Preventable/Avoidable -Emergent - ED Care Needed - Not Preventable/Avoidable		X	X	X
16	OVERALL	HEDIS; Medicaid Child Core; NQF 1392	Well-Child Visits in the First 15 Months of Life	Percentage of patients who turned 15 months old during the measurement year and who had the following number of well-child visits with a PCP during their first 15 months of life. Seven rates are reported: •No well-child visits •One well-child visit •Two well-child visits •Three well-child visits •Four well-child visits •Five well-child visits •Six or more well-child visits			X	
17	OVERALL		Hospital costs related to avoidable inpatient stays and treat-and-release ED visits				X	

Appendix A2: Crosswalk Between Metrics and Evaluation Hypotheses

Metric Number	Evaluation[1]	Source	Metric Name	Metric Description	Hypothesis			
					Chronic Disease Outcomes		Health Outcomes	Care
					1		2	3
18	OVERALL & DIABETES	HEDIS; Medicaid Adult Core; NQF 0057	Comprehensive Diabetes Care: Hemoglobin A1C Testing	The percentage of members 18-75 years of age with diabetes (type 1 and type 2) who received an HbA1c test during the measurement year.	X		X	
19	OVERALL & DIABETES	HEDIS; NQF 0055	Comprehensive Diabetes Care: Eye Exam	The percentage of members 18-75 years of age with diabetes (type 1 and type 2) who received a retinal or dilated eye exam during the measurement year or a negative retinal or dilated eye exam in the year prior to the measurement year.	X		X	
20	OVERALL		Treat-and-release ED visits	Treat- and -release visits to an emergency department			X	

[1]Metrics will be utilized for the overall evaluation of the DSRIP , the evaluation of hospital projects related to specific chronic conditions (e.g. asthma, cardiac care, diabetes, etc.), or both.

[2]not currently endorsed by NQF

IV. Timeline and Deliverables

Waiver Demonstration Period: 8/1/2017 to 6/30/2022
Demonstration Period for OUD-SUD Initiative: 10/31/2017 to 6/30/2022
Project Period: 1/1/2019-12/31/2023

Deliverables:

Stakeholder Reports
Stakeholders Report on MLTSS: 7/1/2020
DSRIP Stakeholders Report: 9/30/2020
OUD/SUD Program Stakeholders Interview: 7/30/2022

Annual Reports
Annual Report of Metrics for fiscal year 2017-2018: 10/31/2019
Annual Report of Metrics for fiscal year 2018-2019: 7/30/2020
Annual Report of Metrics for fiscal years 2020-2021: 7/30/2022

Note: OUD-SUD metrics will not be part of annual reports.

Interim and Final Evaluation Reports
Draft Interim Evaluation Reports (non-DSRIP components): 6/30/2021
DSRIP Final Evaluation Report: 12/15/2021
Draft Final Evaluation Reports (non-DSRIP components): 9/30/2023

Note: The evaluation reports for the OUD-SUD initiative will be separate from the other components.

Finals due 60 days after receiving CMS comments on Draft Evaluation

V. Faculty Bios

Sujoy Chakravarty, PhD (Principal Investigator), Assistant Research Professor and Health Economist at the Rutgers Center for State Health Policy (CSHP), will direct all aspects of the project including model conceptualization, design and analysis. Dr. Chakravarty led the evaluation of the 2012-2017 NJ Medicaid 1115 Comprehensive Waiver Demonstration that included analyses of the MLTSS and DSRIP programs among other reforms. Dr. Chakravarty has considerable expertise in Medicaid policies and their potential effects on healthcare services and outcomes and is an expert in policy evaluation design and analysis strategies. The evaluation involved examining the effect of several simultaneous policy changes relating to eligibility, financing and population health management on specific waiver populations by analyzing Medicaid fee-for-service claims and managed care encounter data. He has published several papers and reports utilizing econometric techniques such as panel data estimation and difference-in-differences modelling to examine provider services, healthcare utilization, prescription coverage, and racial and ethnic disparities in access.

Joel C. Cantor, ScD (Senior Research Advisor), Distinguished Professor of Public Policy and CSHP Director will work closely with Dr. Chakravarty to ensure that the study design and project findings are relevant to policymakers and stakeholders. Dr. Cantor has a deep understanding of the New Jersey policy and health care delivery context and is an expert in the communication of research findings to policy and practice audiences. He is a member of the National Advisory Committee of the AcademyHealth Translation and Dissemination Institute, and has great depth of experience in conducting policy studies and engaging with policy audiences. Dr. Cantor is the founding (1999) director of Rutgers Center for State Health Policy, where he has led policy-engaged research for over two decades focusing on healthcare financing, regulation and delivery, primarily at the state level. A substantial body of his work focuses on Medicaid, where he has led quantitative and mixed-methods work related to evaluating the impact of federal and state policies.

Laura Pizzi, PharmD, MPH (Co-Investigator), will lead the project's cost-effectiveness analysis. She is Professor and Director of the Center for Health Outcomes, Policy, and Economics at Rutgers University. Her research focuses on the economic analysis of healthcare interventions and new models of delivering care. Most of her research during the past 20 years has focused on the cost effectiveness of healthcare interventions for the prevention and treatment of chronic diseases. Dr. Pizzi has authored or co-authored more than 75 peer-reviewed articles, is Deputy Editor of *American Health and Drug Benefits,* editorial board member for *PharmacoEconomics*, and is co-editor of the text *Economic Evaluation in U.S. Healthcare: Principles and Applications.*

Enclosure

cc: Francis McCullough, Director, Division of Medicaid Field Operations - East
Ricardo Holligan, Deputy Director, Division of Medicaid Field Operations - East