DEPARTMENT OF HEALTH & HUMAN SERVICES
Centers for Medicare & Medicaid Services
7500 Security Boulevard, Mail Stop S2-01-16
Baltimore, Maryland 21244-1850

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

State Demonstrations Group MAR 0 8 2017

Deborah Fournier
Director
Office of Medicaid Business and Policy
New Hampshire Department of Health and Human Services
129 Pleasant Street
Concord, NH 03301-6521

Dear Ms. Fournier:

The Centers for Medicare & Medicaid Services (CMS) has reviewed New Hampshire's proposed evaluation design for the 1115(a) demonstration (Project No. 11-W-00301/1), entitled "Building Capacity for Transformation" (BCT).

Please see the attached document, which details CMS' comments and feedback on the proposed evaluation design. In accordance with the Special Terms and Conditions (STC), please review these comments and incorporate the necessary changes into the quality strategy design for submission to CMS within 60 business days of the date of this letter.

If you have any questions, please do not hesitate to contact your project officer, Mr. Adam Goldman. Mr. Goldman can be reached at (410) 786-2242, or at Adam.Goldman@cms.hhs.gov.

We look forward to continuing to partner with you and your staff on the New Hampshire BCT demonstration.

Sincerely,

[redacted]

Angela D. Garner
Director
Division of System Reform Demonstrations

Enclosure

cc: Richard McGreal, Associate Regional Administrator, CMS Boston Regional Office

TO: Deborah Fournier and Kelley Capuchino

FROM: CMS

DATE: March 8, 2017

SUBJECT: Comments on New Hampshire's DSRIP Evaluation Plan

On October 17, 2016, the Centers for Medicare & Medicaid Services (CMS) received the draft evaluation design plan for New Hampshire's ("state") Delivery System Reform Incentive Payment (DSRIP) program entitled, "Building Capacity for Transformation" (Project No. 11-W-00301/1). The special terms and conditions (STC) for the state's current section 1115 require it to develop an evaluation design for approval by CMS. The STCs define specific core components, described below, that must be included in the design. CMS reviewed the report against those requirements and examined the strengths and limitations of the proposed design.

The following is a summary of the state's evaluation plan, an analysis if the degree to which it aligns with the STCs, and recommendations for strengthening the clarity and rigor of the evaluation design.

A. Overview of the state's evaluation design

The state's DSRIP program aims to transform the way physical and behavioral health care is delivered to Medicaid beneficiaries with behavioral health conditions, substance use disorders, and/or substance misuse (henceforth referred to as behavioral health conditions). To accomplish this, the state is developing and maintaining seven regional Integrated Delivery Networks (IDN) that serve Medicaid beneficiaries with behavioral health needs. Each IDN will be eligible to receive incentive payments based on implementation of six projects that address the needs of this population. To collect the payments, the IDNs must achieve milestones and report or improve performance on quantitative metrics associated with each project. The projects focus on four categories: (1) bolstering appropriate behavioral health capacity, (2) promoting integration of physical and behavioral health providers through physical or virtual integration, (3) promoting smooth transitions across the continuum of care, and (4) ensuring IDN participation in alternative payment models (APM).

The purpose of the evaluation is to assess the extent to which the state's DSRIP achieved the goals of improved physical and behavioral health treatment capacity, improved integration of physical and behavioral care, improved transition across settings, movement toward IDN adoption of APMs, and reduced cost of care for Medicaid beneficiaries with behavioral health conditions.

The evaluation will rely on both quantitative and qualitative research methods. The quantitative portion of the study will use both cross-sectional and sequential, cross-sectional analyses. The cross-sectional analysis will compare outcomes of interest at a single point in time

between beneficiaries with and without behavioral health conditions who are served by an IDN. The sequential, cross-sectional analysis will use a pre-post design, annually comparing the outcomes of interest one year before the DSRIP start date to outcomes in the years following implementation. Secondary analyses will include multiple regression analysis to control for member and IDN characteristics.

The quantitative analyses will use a variety of data sources, including Medicaid claims and encounter data from the state's Medicaid Management Information System (MMIS), IDN electronic health record data, discharge data from the New Hampshire Hospital (an inpatient psychiatric facility), and state-computed Healthcare Effectiveness Data and Information Set (HEDIS) measures. In addition, the evaluation will use the Comprehensive Health Care Information System, the state's all-payer claims database, for data on beneficiaries covered under the state's premium assistance program.

For the qualitative analysis, the evaluation will include surveys with Medicaid beneficiaries, IDN administrators, health care and community-based providers, and health information technology (IT) stakeholders. The surveys will focus on improvements in care coordination and integration, perceptions of the IDNs, health IT, and enhancements to the IT system. Finally, the independent evaluator will conduct semi-structured interviews with the same four groups of stakeholders as the surveys. The interview topics will include experiences with health care, experiences with care coordination and integration, and perceptions of the IT systems.

B. Review of the evaluation design's alignment with STCs

Although the state's evaluation design was generally responsive to the STCs, there are several important discrepancies between the two (see Table 1 for a detailed comparison of the evaluation design requirements, per the STCs, and the evaluation plan submitted by the state).

Table 1. Comparison of evaluation design requirements and New Hampshire's evaluation design plan

Requirements specified in STCs	Requirements addressed in report	Requirements not addressed in report
The draft evaluation design must include a discussion of the goals, objectives, and evaluation questions specific to the entire delivery system reform demonstration.	• Included discussion of demonstration goals and objectives specific to the entire delivery system reform demonstration • Included evaluation questions specific to quality of care, cost of care, avoidable re-hospitalization, and access to mental health care	• Did not include evaluation questions that assess the demonstration goals and objectives related to workforce development and APMs
The draft design must discuss the outcome measures that will be used in evaluating the impact of the demonstration during the period of approval, particularly among the target population, specific testable hypothesis, including those that focus on target populations for the demonstration and more generally on beneficiaries, providers, plans, market areas, and public expenditures.	• Included discussion of outcome measures • Included discussion of specific testable hypotheses that focus on the target populations	• Did not discuss testable hypotheses that focus on providers, plans, or market areas

Requirements specified in STCs	Requirements addressed in report	Requirements not addressed in report
The draft design must discuss the data sources, including the use of Medicaid encounter data, and sampling methodology for assessing these outcomes.	• Included discussion of data sources and sampling methodology for stakeholder surveys and semi-structured interviews	• Did not discuss sampling methodology for the treatment and comparison groups proposed for the quantitative analyses (e.g., for outcomes measures that rely on data from electronic health records)
The draft evaluation design plan must describe how the effects of the demonstration shall be isolated from other initiatives occurring within the state.	• None	• Did not describe how the effects of the demonstration will be isolated from other initiatives occurring within the state
The demonstration evaluation will meet all standards of rigor, including the use of best available data; controls for and reporting of the limitations of data and their effects on results; and the generalizability of their results.	• Included several quantitative and qualitative data sources • Mentioned that the evaluation will use controls and adjustments where appropriate and available • Discussed potential data limitations and their effects on results • Mentioned that the evaluation will discuss the generalizability of results in the context of the limitations	• None
The state will test the following hypotheses in its evaluation demonstration: • Individuals with co-occurring physical and behavioral health issues will receive higher quality of care after IDNs are operating • The total cost of care will be lower for Medicaid beneficiaries with co-occurring physical and behavioral health issues after IDNs are operational • The rate of avoidable re-hospitalizations for individuals with co-occurring physical and behavioral health issues will be lower at the end of the demonstration than prior to the demonstration • Percentage of Medicaid beneficiaries waiting for inpatient psychiatric care will be lower at the end of the demonstration than prior to the demonstration • Average wait times for outpatient appointments at community mental health centers will be lower at the end of the demonstration than prior to the demonstration	• Included hypotheses that test changes in the quality of care, cost of care, avoidable re-hospitalization, percentage of Medicaid beneficiaries waiting for inpatient psychiatric care, and average wait times for outpatient appointments at community mental health centers at the end of the demonstration • Hypotheses presented in Appendix A suggest that the evaluation will focus on individuals with co-occurring physical and behavioral health conditions, but only a few quality-of-care measures proposed focus on these individuals	• Hypotheses presented on page 9 did not address Medicaid beneficiaries with co-occurring physical and behavioral health conditions

Requirements specified in STCs	Requirements addressed in report	Requirements not addressed in report
The evaluation design must, at a minimum, address the following research questions: • 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? • To what extent has the DSRIP enhanced the state's health IT ecosystem to support delivery system and payment reform? Has it specifically enhanced these four key areas through the IDNs: governance, financing, policy/legal issues and business operations? • To what extent has the DSRIP improved integration and coordination between providers, including bi-directional integrated delivery of physical, behavioral health services, SUD services, transitional care, and alignment of care coordination and to serve the whole person?	• Examined whether there is a statistically significant improvement in measures of quality of care and costs after implementation of DSRIP • Addressed the extent to which DSRIP has enhanced the state's health IT ecosystem to support delivery system reform and payment reform through surveys • Addressed the extent to which DSRIP improved integration and coordination between providers through interviews and surveys	• Did not address whether the DSRIP program achieved improved access to care, better health outcomes, or better health for the population • Did not address the degree to which improvements can be attributed to the activities undertaken under DSRIP • Did not address whether health IT has enhanced governance, financing, policy/legal issues, and business operation, but this can likely be addressed in the key informant surveys
The state must select a preferred research plan for the applicable research question and provide a rationale for its selection. To the extent applicable, the following items must be specified for each design option that is proposed: • Quantitative or qualitative outcome measures • Baseline and/or control comparisons • Process and improvement outcome measures and specifications • Data sources and collection frequency • Robust sampling designs (e.g., controlled before-and-after studies, interrupted time-series design, and comparison group analyses) • Cost estimates • Timelines for deliverables	• Included quantitative and qualitative outcome measures • Included baseline and control comparison • Included process and improvement outcome measures and specifications • Included data sources and collection frequencies • Mentioned beneficiaries who do not have any indicator of behavioral health conditions as a potential comparison group for one measure (Measure 1.1.17—adolescent well-care visit) • Included timeline for deliverables	• Relied on pre-post comparisons for quantitative analysis measured on an annual basis, with one year of pre-demonstration data; however, one year of data does not provide a robust comparison • Indicated that cost estimates for the evaluation are currently unavailable and will be determined through the competitive bid process
The evaluation designs proposed for each question may include analysis at the beneficiary, provider, and aggregate program level, as appropriate, and include population stratifications to the extent feasible, for further depth and to glean potential non-equivalent effects on different subgroups. In its review of the draft evaluation plan, CMS reserves the right to request additional levels of analysis.	• Included analysis at the beneficiary level • Included stratification by age for certain measures • Included stratification by type of disorder (e.g., schizophrenia and bipolar disorder)	• None

First, the STCs specify that the draft design must include evaluation questions specific to the objectives of the entire delivery system reform demonstration. These objectives include workforce development, access, technology, incentives, recovery models, integration, care transitions, and APMs. Currently, the draft evaluation plan does not include evaluation questions focused on workforce development and APMs, nor does it assess the impact of DSRIP on these outcomes. To address workforce development, the state could analyze secondary data from the Area Health Resources File[1] or the Bureau of Labor Statistics.[2] Regarding APMs, the evaluator may be able to identify data sources that assess provider-level health IT infrastructure as a way to indirectly examine APM readiness (for instance, Medicare Electronic Health Record Incentive Program Eligible Professionals Public Use File). Alternatively, the evaluator might include questions on these topics in provider and IDN administrator surveys.

The STCs also require that the evaluation design address the following research questions:

1. 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?
2. To what extent has the DSRIP enhanced the state's health IT ecosystem to support delivery system and payment reform? Has it specifically enhanced these four key areas through the IDNs: governance, financing, policy/legal issues, and business operations?
3. To what extent has the DSRIP improved integration and coordination between providers, including bi-directional integrated delivery of physical, behavioral health services, SUD services, transitional care, and alignment of care coordination and to serve the whole person?

Currently, the evaluation plan does not fully address how each of these questions will be answered. For example, it does not describe how it will assess whether DSRIP achieved better health outcomes or population health. The plan proposes to use IDN electronic health records as a data source, which provides an opportunity to assess the impact of DSRIP on health outcomes for those seeking care within an IDN. Another option is to include carefully targeted questions regarding health status or health-related quality of life in stakeholder surveys administered to Medicaid beneficiaries. To assess the impact of DSRIP on population health, the evaluator could turn to additional data sources, such as the state's Behavioral Risk Factor Surveillance System. This state-based system of health surveys monitors several measures of health in the state, including health status, health-related quality of life, and chronic health conditions.

The evaluation plan also does not address how the effects of the demonstration will be isolated from other initiatives in the state. For instance, the state received a design cooperative agreement for a state innovation model (SIM) to improve the quality and efficiency of health

---

[1] See http://ahrf.hrsa.gov/.

[2] See http://www.bls.gov/oes/current/oes_nh.htm#29-0000

care delivery, expand the use of health IT, and lead improvement programs focused on population health. Activities carried out under the SIM agreement align with many of the activities being carried out through the state's DSRIP program.

Although the evaluator will not be able to determine conclusively whether any effects can be attributed to DSRIP, several methods could be used to support potential findings, such as falsification tests. For instance, the evaluator could assess changes in outcomes of interest over the same time period for a population that CMS would not expect to be affected by DSRIP. More specifically, the evaluator could focus on physical health outcomes (for instance, cervical, breast, or colorectal cancer screening; cholesterol screening; emergency department [ED] visits; preventable ED visits; and costs of care) for beneficiaries without behavioral health conditions, expanding the use of this comparison group for the evaluation.[3] The evaluator could also assess trends in outcomes that CMS would not expect to be affected by DSRIP, such as non-preventable ED visits.

The STCs specify that the evaluation design should test whether individuals with co-occurring physical and behavioral health conditions will have higher quality care, lower total costs, and reduced rates of avoidable re-hospitalizations after the IDNs are operating. The design does include three quality-of-care measures that focus on individuals with co-occurring physical and behavioral health conditions (diabetes screening for people with schizophrenia or bipolar disorder who are using antipsychotic medication, diabetes screening for people with diabetes and schizophrenia, and cardiovascular monitoring for people with cardiovascular disease and schizophrenia). However, the plan does not assess whether DSRIP influences total costs or avoidable re-hospitalizations specifically for individuals with co-occurring physical and behavioral health conditions.

CMS suggests that the state update its selection strategy for the study population (the population exposed to the DSRIP demonstration or the treatment group) to also identify a subpopulation with physical health conditions that commonly co-occur in people with behavioral health conditions. At a minimum, the state should identify beneficiaries with a behavioral health condition who also have a claim in which the primary or secondary diagnosis is metabolic syndrome, dyslipidemia, diabetes, obesity, hypertension, asthma, or chronic obstructive pulmonary disease. The analyses should then be carried out on both populations: (1) all beneficiaries with a behavioral health condition and (2) beneficiaries with co-occurring physical and behavioral health conditions.

C. Recommendations to improve the clarity and rigor of the evaluation design

Besides addressing the discrepancies between the plan and the STCs, CMS suggests that the state take several steps to improve the clarity of the evaluation plan and the rigor of the design.

---

[3] The evaluation plan only proposes to compare these two groups for the adolescent well-care measure.

1. Improving clarity

The evaluation plan lays out several research questions, hypotheses, and outcomes that, at times, are not fully aligned. For instance, one key question focuses on the state's health IT ecosystem, but the hypotheses presented do not address this question. The appendix clarifies the measures that will be used to examine this research question, but it would be helpful to see this information laid out in the body of the document. To ensure all aspects of the design are aligned, CMS recommends that the state restructure its plan so that each research question is immediately followed by hypotheses that address the question (when appropriate) and a brief list of measures to address the research question and hypotheses. This information currently appears in various places throughout the plan, but restructuring the plan as CMS suggests will ensure the questions, hypotheses, and measures are all linked. For example, the state could lay out the plan in this way:

Research question 1: Was the DSRIP program effective in achieving the goal of lower cost of care?

Hypothesis 1.1: The total cost of care will be lower for Medicaid beneficiaries with behavioral health needs after IDNs are operating.

Outcomes measures: Total cost of all care; total cost of all inpatient care; total cost of all outpatient care; total cost of ED care; total cost of behavioral health care; total cost of outpatient behavioral health care; total cost of inpatient behavioral health care; total cost of ED behavioral health care

CMS also suggests that the state include additional detail about the quantitative methods proposed in the evaluation design. The evaluation plan states that nonparametric tests will be the primary method of analysis, and regression will be the secondary method. The state does not justify its use of nonparametric tests over regression analysis. This information would be helpful given that regression analysis is typically a more rigorous approach than the nonparametric tests proposed. Further, the appendix does not include regression analysis as a comparison method for any quantitative outcomes measures. Additional information about when regression analysis will be applied would be helpful.

CMS recommends that the state provide more detail about how it will select the pre-demonstration comparison group. The evaluation plan says comparisons will be made between the pre-demonstration period and the demonstration period for the study population, but the study population includes Medicaid beneficiaries who have a behavioral health condition and are served by an IDN during the study period. Because IDNs are being developed through the DSRIP demonstration, the state will not be able to identify beneficiaries in IDNs in the pre-period. CMS suggests that the state further explain its inclusion and exclusion criteria for the pre-period comparison group.

Lastly, the evaluation design does not discuss how costs of care will be measured for beneficiaries who are covered under managed care. This is particularly important because almost all Medicaid beneficiaries in the state are covered by a managed care organization (MCO). Total cost of care to the state can be estimated from the capitated payment rates paid to MCOs, but this information cannot be used to determine the cost of the specific service types included in the

evaluation plan (for instance, cost of psychiatric care). Some managed care organizations report the costs of care on the encounter claims records that they submit to state Medicaid agencies, particularly when the plans do not sub-capitate provider payments. Evaluations of managed care systems have also developed approaches to assigning costs to the encounter claims records that managed care plans generate, frequently known as shadow prices. CMS recommends that the evaluation plan describe how the costs of care will be estimated for purposes of the evaluation.

2. Increasing rigor

The evaluation plan can also be adapted in several ways to improve the rigor of the design. First, the state could use an interrupted time-series design, a more rigorous econometric method for estimating changes in trends than nonparametric tests and multivariate regression. This design requires multiple observations from the pre-demonstration period to identify pre-demonstration trends in outcomes. CMS believes the evaluation of the state's demonstration should account for pre-demonstration trends in outcomes—partly because of all the other health care initiatives occurring in Medicaid and Medicare programs, which are likely to affect many of the state's proposed measures. Therefore, CMS suggests that the evaluation include at least two years of data before the demonstration, measured quarterly instead of annually. The trend in eight quarters of data from the pre-demonstration period can be used to estimate trends in the outcomes of interest in the post-demonstration period. The predicted trends would then be compared to the trends that occurred during the demonstration to determine if statistically significant changes in trends occurred after implementation.

CMS also recommends considering an alternative comparison group. The evaluation plan considers Medicaid beneficiaries who do not have any indicator of behavioral health conditions as a comparison group for certain measures. As described above, CMS believes the outcomes of such beneficiaries can be useful for a falsification test, but these individuals may not be an appropriate comparison group because they are fundamentally different from the treatment group. A potential alternative comparison group is Medicaid beneficiaries receiving care through the NH Health Protection Premium Assistance section 1115 demonstration, a population excluded from the demonstration.[4] At this point, the extent to which this population will be affected by the demonstration is unclear. If premium assistance beneficiaries are not attributed to IDNs, they may act as a useful comparison group.

Further, CMS believes more details on how the pre-demonstration comparison group will be identified and selected for the evaluation is warranted. This comparison group will serve as the primary source of information for what would have occurred had the DSRIP demonstration not been implemented and how this comparison group is defined and identified in the data will be an important factor in the reliability of the estimated effects of this demonstration.

CMS also expects that the demonstration effects may vary by IDN. The evaluation plan currently assesses the outcomes of interest at the state level, but it does not include analyses by

[4] Per STC number 19, individuals served under the New Hampshire Health Protection Program Premium Assistance section 1115 demonstration (11-W-00298/1) are excluded from the DSRIP demonstration.

IDN. CMS recommends including additional analyses comparing trends in the outcomes of interest across IDNs to identify variation in outcomes.

More broadly, the current evaluation design does not adequately address the variation in service needs across Medicaid beneficiaries with behavioral health conditions. Beneficiaries with different behavioral health needs will likely be affected differently by the DSRIP demonstration, in part because the IDNs are likely to implement different models of integrated care for different subgroups. For example, primary care is increasingly acting as a gateway to mental health services for beneficiaries with mild to moderate mental illness (Unützer et al. 2013). Consequently, integrating mental health services into primary care settings is an effective model for this population. On the other hand, people with SMI are significantly more likely to have physical health conditions than the general US population (Scharf et al. 2014), yet they are more likely to seek treatment in mental health settings than in primary care (Unützer et al. 2013). As a result, reverse integration, or integrating primary care into behavioral health settings, can be an effective model for these higher-need beneficiaries. To address these differences, CMS recommends that the evaluator consider running separate models for certain subpopulations defined by their clinical characteristics. These models should focus on specific outcomes of interest. For instance, the evaluators could assess whether Medicaid beneficiaries with SMI have better access to physical health care after IDNs become operational.

Overall, the state's DSRIP evaluation plan has many positive features, including plans to incorporate data from a variety of sources, both before and after the demonstration begins, and cross-sectional approaches to identifying the effect of the IDNs. However, the design does not fully comply with the STCs, and several changes could be made to strengthen the clarity and rigor of the evaluation. At a minimum, the state should revise the evaluation plan to address he requirements identified in Table 1 above—including, but not limited to, how it will assess behavioral health workforce development, health outcomes, and progress toward APMs; how DSRIP effects will be isolated from other activities occurring in the state; and additional subgroup analyses to assess the impact of DSRIP on people with co-occurring physical and behavioral health conditions. The state should also consider several additional revisions to improve the clarity and rigor of the evaluation plan.

References

Scharf, D.M., N.K. Eberhart, N.S. Hackbarth, M. Horvitz-Lennon, R. Beckman, B. Han, S.L. Lovejoy, H.A. Pincus, and M.A. Burnam. "Evaluation of the SAMHSA Primary and Behavioral Health Care Integration (PBHCI) Grant Program: Final Report (Task 13)." Santa Monica, CA: RAND Corporation, 2014. Available at http://www.integration.samhsa.gov/integrated-care-models/RAND_Evaluation_SAMHSA_PBHCI_Grant_Program.pdf. Accessed November 29, 2015.

Unützer, J., H. Harbin, M. Schoenbaum, and B. Druss. "The Collaborative Care Model: An Approach for Integrating Physical and Mental Health Care in Medicaid Health Homes." May 2013. Available at https://www.medicaid.gov/State-Resource-Center/Medicaid-State-Technical-Assistance/Health-Homes-Technical-Assistance/Downloads/HH-IRC-Collaborative-5-13.pdf. Accessed November 29, 2015.