New York State Department of Health

Delivery System Reform Incentive Payment (DSRIP) Evaluation Design

22. Evaluation Design. The Evaluation Design shall include the following core components to be approved by CMS:

- ***a) Research questions and hypotheses: This includes a statement of the specific research questions and testable hypotheses that address the goals of the demonstration, including:***
  - ***i. safety net system transformation at both the system and state level;***
  - ***ii. accountability for reducing avoidable hospital use and improvements in other health an public health measures at both the system and state level and***
  - ***iii. efforts to ensure sustainability of transformation of/in the managed care environment at the state level.***

The research questions will be examined using appropriate comparison groups and studied in a time series.

Overarching Research Questions (P.7, Request for Proposals)

1. To what extent did PPSs achieve health care system transformation?
2. Did health care quality improve as a result of clinical improvements in the treatment of selected diseases and conditions?
3. Did population health improve as a result of implementation of the DSRIP initiative?
4. Did utilization of behavioral health care services increase as a result of DSRIP?
5. Was avoidable hospital use reduced as a result of DSRIP?
6. Did DSRIP reduce health care costs?
7. What were the successes and challenges with respect to PPS planning, implementation, operation and plans for program sustainability from the perspectives of DSRIP planners, administrators and providers, and why were they successful and challenging?

<u>Time Series Analysis</u>

We propose a mixed methods strategy to meet the project objectives throughout the evaluation. This strategy offsets the weaknesses inherent in single method approaches and allows us to confirm, cross-validate, and corroborate the findings.

The time series part of the project will emphasize comparison of health care service delivery, health improvements, and cost to the Medicaid program at the state level over the study period. It will also do an inter-PPS analysis to identify components that posed particular successes or challenges for implementation and outcomes by Difference in Differences analysis.

The time series analysis will focus on the seven research questions with the following outcomes and conditions, as appropriate:

<u>Research Question 1</u>. To what extent did Performing Provider Systems achieve health care system transformation, including increasing the availability of behavioral health care?

1. Integration of service delivery.
2. Care Coordination and Connecting Settings
3. Availability and Use of Primary Care and Medicaid Primary Care Spending
4. Availability and use of behavioral health services, and behavioral health care spending.
5. Medicaid spending on emergency department and inpatient services.
6. Access to and utilization of primary and preventative services by uninsured, non-utilizing, and low-utilizing populations**.**

Research Question 2. Did health care quality improve as a result of clinical improvements in the treatment of selected diseases and conditions?

7. Behavioral Health (required)
8. Cardiovascular Health
9. Diabetes Care
10. Asthma
11. HIV/AIDS
12. Perinatal Care
13. Palliative Care
14. Renal Care

Research Question 3. Did population health improve as a result of implementation of the DSRIP initiative?

15. Promote Mental Health and Prevent Substance Abuse (MHSA)
16. Prevent Chronic Diseases
17. Prevent HIV and STDs
18. Promote Healthy Women, Infants and Children

Also, racial and ethnic disparities will be addressed with respect to the following metrics: premature deaths, newly diagnosed cases of HIV, preterm births, adolescent pregnancy rate per 1,000 females aged 15-17, percentage of unintended pregnancy among live births, and infants exclusively breastfed while in the hospital. Disparities on these outcomes will be measured as ratios and will be treated as additional outcomes at the statewide level with the prediction that these ratios will show improvement (i.e., will be reduced) following DSRIP implementation.

Research Question 4. Was Avoidable Hospital Use Reduced as a Result of DSRIP?

Research Question 5. Did DSRIP reduce health care costs?

In all these 20 cases, it is hypothesized that following the introduction of DSRIP, the health care of the Medicaid patients have become better and also the program has become economically more efficient. Due to small sample size and multiple hypotheses testing, correct significance levels have to be determined by controlling the false discovery rate (FDR), rather by conventional Bonferroni bounds.

The following two additional questions are also to be answered using the available data:

Research Question 6. Was DSRIP cost effective in terms of New York State and Federal governments receiving adequate value for their investments?

Research Questions 7. What were the successes and challenges with respect to PPS planning, implementation, operation, and plans for program sustainability from the perspectives of DSRIP planners, administrators, and providers, and why were they successful or challenging?

Qualitative Analysis

Process/Implementation Study
Qualitative data will provide context for the quantitative data assessing the overall evaluation's research questions 1-4, which focus on system transformation, clinical improvement, and population wide projects(domains 2-4). These questions focus on the implementation of projects initiated with the DSRIP program. Qualitative data will also address the 7th research question, which asks about successes and challenges related to different aspects of the DSRIP program.

Additionally, the implementation study will address the following research questions:

Facilitators and Barriers to Pay-for-Performance Metrics

1. What services are being provided in each project dimension?
2. What are the most critical components of each project?
3. Have the selected projects been implemented as designed/intended (e.g., modifications or adaptions, consistency with program design, fidelity to a model?)
4. How well does the program connect with other programs and services received by participants?
5. What are the key factors in the project's environment (e.g., the larger community, the network of services, community based organizations) that influence project implementation?
6. What barriers or challenges been encountered during service delivery?
7. What strategies have been utilized? What were there outcomes?
8. How have other health care initiatives impacted DSRIP?
9. How satisfied are DSRIP stakeholders with program planning?
10. How satisfied are DSRIP stakeholders with program implementation and operation?

Perceived Outcomes

1. What changes have there been to health care system overall?
2. What change shave there been behavioral health care?
3. What changes has there been to population health?
4. How effective do DSRIP stakeholders perceive the projects to be? Perceive DSRIP to be overall?
5. Which participants seem to be benefiting the most and the least? Why?
6. What recommendations are offered regarding DSRIP improvement?

Patient Experience

1. How has the patient experience changed?
2. How satisfied are patients with the change?

b) The design will include a description of the quantitative and qualitative study design (e.g., cohort, controlled before-and-after studies, interrupted time series, case-control, etc.), including a rationale for the design selected. The discussion will include a proposed baseline and approach to comparison. The discussion will include approach to benchmarking, and should consider applicability of national and state standards. The application of sensitivity analyses as appropriate shall be considered.

Time Series Analysis

This part of the project will emphasize comparison of health care service delivery, health improvements, and cost to the Medicaid program at the state level over the study period. It will also do an inter-PPS analysis to identify components that posed particular successes or challenges for implementation and outcomes by DID analysis. Possible improvement in

twelve broad categories of health care under four domains is envisioned due to DSRIP program.

The work involves using the interrupted time series/segmented regression on the following twenty statewide time series to examine if post DSRIP values are better that those of the pre-DSRIP period from the standpoint of efficiency of the newly designed Medicaid program. These twenty series are listed under the following five broad questions. The available data sources include information from Medicaid Claims, Medicare Claims, Statewide Planning and Research Cooperative System (SPARCS), Minimum Data Set (MDS), Consumer Assessment of Healthcare Providers and Systems (CAHPS®), New York Vital Statistics, Expanded Behavioral Risk Factor Surveillance System (eBRFSS), New York HIV/AIDS Case Surveillance Registry, Uniform Assessment System (UAS), US Census and American Community Survey.

Even though we will use the interrupted time series design as the workhorse of our analysis, we will also explore if an appropriate non-DSRIP control group can be identified for Differnce in Differences (DID) analysis in our time series analysis and panel data can be developed for some outcome variables. We realize a non-Medicaid population as a control group will be hard to identify because it would likely to differ in many ways from Medicaid population including demographically, socioeconomically and, more importantly, by health. However, the DID estimator only requires that in the absence of the treatment, the average outcomes of the treated and control groups would have followed parallel paths over time. Even this assumption may not be reasonable in our context because the pre-treatment characteristics may be associated with the dynamics of the outcome variable that can affect the control and the treatment groups asymmetrically. In this situation we will experiment with Abadie (2005)'s simple two-step semiparametric strategy to estimate the average treatment effect of the treated. These methods will have to be corrected for serial correlation in the outcome variable by using appropriate cluster analysis.

Qualitative Analysis

Process/Implementation Study Design

A mixed method (quantitative and qualitative) process evaluation will be conducted. A mixed method evaluation was chosen because utilizing several methods of data collection offsets the weaknesses inherent in single method approaches and allows researchers to confirm and cross-validate key findings[i].

Researchers will use three major data methods from a number of relevant stakeholders: **Interviews** with PPS administrators, **focus groups** with project-associated providers, and **surveys** with patients and project-associated providers. These data sources will be used to collect data on three major focal points: the DSRIP program overall, individual projects, and patient experience. In general, interviews and focus groups will be the major data source for overall DSRIP program data, surveys of patients will be the major source of patient satisfaction and experience, and surveys of providers will be the major source of project specific data. In the final stage of the analysis, findings from the different data types and sources (quantitative and qualitative) will be triangulated to develop an integrated analysis.

Each data source is described below:

Focus Groups
Focus groups will be conducted with select project-associated providers. Drawing from research on best practices for conducting focus groups, the number of participants for each focus group will be limited to 10-12 individuals; this group size allows participants sufficient time to share insights, yet is large enough to provide a diversity of perspectives. Focus groups will be conducted with each PPS separately. The focus groups will be guided by a focus group protocol, with questions tailored to each PPS group. Each focus group will last approximately 1-1.5 hours. Focus group participants will be informed of the research protocol regarding confidentiality before the session begins. This includes reporting the findings as a group and not associating anyone with individual remarks. With the permission of the participants, all qualitative focus groups and interviews will be audio-recorded and transcribed verbatim, and field notes will be taken to document the process. Focus groups will be held with associated providers and administrators at least once per PPS.

Interviews
Key informant interviews will be conducted with PPS staff and administrators annually. Interviews will be scheduled at the convenience of the PPS staff and administration, and will be conducted either by phone or in person. The interviews will be guided by a semi-structured interview protocol, and should take about an hour to complete. A core set of questions will be asked of all key informants, and a subset of questions and probes will be developed based on each key informant's roles, knowledge, and responsibilities.

Survey for Patients
An electronic survey will be sent to a representative sample of patients from each PPS to gauge patient satisfaction and experience. The survey will be sent electronically and will include a combination of closed and open-ended questions, which will be aligned with questions asked to other stakeholders in the qualitative data collection procedures. The survey will be designed using best practices and in a manner that reduces burden on respondents (Dillman et al., 2014).

Survey for Project Providers
In order to collect uniform information on the functioning of individual projects, an electronic survey will be administered at regular intervals to project-associated providers. PPS administrators will provide lists of providers who are associated with each of their projects, and these providers will comprise the sample. This survey will focus specifically on progress within individual projects, as well as barriers and facilitators to project implementation, and perceived effectiveness. This survey will provide user-based feedback which will aid in the providing individualized feedback to each PPS for quality improvement of their projects (Bate & Robert, 2007).

Table 1: Summary of Data Sources and Areas of Inquiry

Areas of Inquiry	Interviews with PPS admin/ planners	Focus Groups with providers	Surveys with Patients	Surveys with Providers on projects
DSRIP Program Overall				
Program planning, operation, and effectiveness	X	X		X
Program outcomes and challenges	X	X		X
Plans for program sustainability	X			X
Effectiveness of governance structure and provider linkages	X	X		
Facilitators and barriers to PPS's achieving progress on pay-for-reporting/pay-for-performance metrics	X	X		X
Contractual and financial arrangements	X	X		
Changes in the delivery of patient care	X	X		X
The effect of other ongoing health care initiatives (e.g., New York Prevention Agenda, Affordable Care Act) on DSRIP implementation and operation	X	X		X
Project Specific				
Progress/effectiveness of projects focused on system transformation	X	X		X
Progress/effectiveness of projects focused on behavioral health	X	X		X
Progress/effectiveness of projects focused on clinical improvement and population			X	X
Identify these issues that are characteristic of particular strategies or projects (in terms of PPP/PPR metrics).	X			X
Patient Experience				
Patient satisfaction & experience			X	

Comparative Analysis

We will evaluate the relative effectiveness of DSRIP projects using both quantitative and qualitative approaches. Our approach will apply quantitative techniques to assess relative PPS performance on domain-specific metrics over time, and build upon the qualitative data collection performed by our qualitative team and led by CHSR. Specifically, we will supplement our quantitative analyses of publicly-available data sets by analyzing other primary data, such as: 1) focus groups, 2) semi-structured key informant interviews with PPS administrators and staff, 3) surveys of providers with semi-structured interview follow-up, and 4) surveys with patients, to provide further contextualization of results.

Further, we will develop a compendium of domain projects across all DSRIP PPS that includes information important to the comparative analysis. The compendium will include information on timeline (start and end dates of implementation), planning decisions (changes that occurred prior to implementation or during implementation), fidelity of the intervention to its original intent (ranked low

to high), relative success to internal expectations (low to high), and previous work (was the program new or building upon existing, pre-DSRIP activity). This compendium will allow the comparative analysis team to examine variation between PPS within projects and across domains in a way that will contribute to our understanding of the DSRIP and exploit less apparent differences between the programs and projects to drive analyses. For example, if two projects look the same "on paper" but one is new and one is based upon an existing initiative, we might see differential outcomes (if we are looking at change over time). This section that follows begins with a presentation of our conceptual framework and then details the methodological approach, measures, and data sources needed for evaluation.

The comparative analysis will be designed to address the 7 research questions presented in the Request for Proposals (RFP) (p.6) with specific emphasis on examining the following 5 specific issues from page 7 of the RFP.

1. Where there is variation in the strategies selected per the PPS project requirements described in the STC documents, assess the effect on the pertinent outcome of PPS's having selected a particular strategy. For example, a comparison would be made in the improvement in diabetes care (Domain 2) between PPS's that implement a project to address this issue and PPS's that do not.
2. The relative effectiveness of particular projects intended to produce the same outcome. For example, among PPS's that opt for a strategy to improve asthma care, compare such improvement between those PPS's that chose to implement a project to expand asthma home-based self-management programs to those PPS's that chose alternative projects to improve asthma care.
3. Identification common to those PPSs receiving or not receiving maximum payment based on project valuation.
4. Comparisons between PPS's operating in different regions of New York to identify successes and challenges associated with local resources or procedures.
5. Patient-level comparisons by factors such as age, sex, race, presence of selected chronic conditions, and mental health/substance abuse status to obtain information on variations in service experience and satisfaction under DSRIP, by patient characteristics.

The research aims for comparative analysis are presented, as follows:

- To compare PPS performance on domain-specific metrics for those that did/did not adopt specific DSRIP projects.
- To evaluate the relative effectiveness of specific strategies employed within specific projects.
- To examine contextual factors related to PPS successes and failures in demonstrating improvement in domain-specific metrics.

c) Performance Measures: This includes identification, for each hypothesis, of quantitative and/or qualitative process and/or outcome measures that adequately assess the effectiveness of the Demonstration in terms of cost of services and total costs of care, change in delivery of care from inpatient to outpatient, quality improvement, and transformation of incentive arrangements under managed care. Nationally recognized measures should be used where appropriate. Measures will be clearly stated and described, with the numerator and dominator clearly defined. To the extent possible, the state will incorporate comparisons to national data and/or measure sets. A broad set of metrics will be selected. To the extent possible, metrics will be pulled from nationally recognized metrics such as from the National Quality Forum, Center for Medicare and Medicaid Innovation, meaningful use under HIT, and

the Medicaid Core Adult sets, for which there is sufficient experience and baseline population data to make the metrics a meaningful evaluation of the New York Medicaid system.

Time Series Analysis

We will use a difference-in-differences estimation methodology to examine specific performance measures in the time before and after the implementation of the DSRIP program comparing PPSs involved in specific interventions to those that were not engaged in those projects. This estimation strategy adjusts for time-based variations in outcomes, helping determine program impacts from other phenomena. Moreover, this approach will give us an aggregate understanding as to whether the overall picture has changed for specific domains based on key measures of interest defined in (http://www.health.ny.gov/health_care/medicaid/redesign/dsrip/cms_official_docs.htm) .

This approach will also require the use of risk-adjusted measures. This will be ideal because it would level the playing field in terms of the dual-eligibles & SSI patients as these individuals tend to seek care at distinct locations and are typically-high utilizers of care. Also, prior to carrying out this analysis, we will endeavor to identify patients and providers (hospitals and medical groups) who were not involved in any DSRIP PPS and understand the trends in use, quality, and spending over time in a separate difference-in-differences analysis.

Qualitative Analysis

Survey items will be selected using existing measures whenever possible to ensure psychometrically rigorous measures are employed.

Outcomes of Interest

Quantitative and qualitative data will be used to aid in the understanding of several outcomes of interest. Outcomes of interest are based on the research questions presented in the request for proposals. Quantitative and qualitative measures will be derived from different sources (e.g., qualitative data is based on analysis of patterns and responses via Atlas-TI).

Table 2. Outcomes and Associated Methods/Data

Outcome	Data
Quantitative	
Avoidable hospital use	3M, AHRQ, Medicaid Claims
Health care cost	Change in spending over time from Medicaid claims, compared to national Medicaid spending growth trend
Qualitative	
PPSs' achievement of health care transformation	Interviews with administrators, focus groups with providers, surveys with providers
Health care quality improvement	Interviews with administrators, focus groups with providers, surveys with providers
Population health improvement	Interviews with administrators, focus groups with providers, surveys with providers
Use of behavioral health care services	Interviews with administrators, focus groups with providers, surveys with providers
Successes and challenges of planning, implementation, and operation	Interviews with administrators, focus groups with providers, surveys with providers, surveys with patients

Comparative Analysis

Measures of Interest. To ground our comparison of PPSs, we have identified a number of measures from Attachment J (NY DSRIP Strategies Menu and Metrics) that have broad-ranging implications on the overall success of the NY DSRIP program. These measures were chosen based on their potential relevance to overall DSRIP goals (e.g., reducing avoidable hospital use by 25% over 5 years) and the four most notable disease areas based on DSRIP project selections and the overall burden of disease in NYS. We will use these metrics as the basis for our comparative analysis of PPSs. Additional metrics can be added based upon priorities of the NYSDOH and project resources. Table 2 provides further detail on selected metrics:

Table 3. Measures of Interest by Domain and Category

Domain/Category	Measure Name	Measure* Steward	Data Source*
Domain 2, A	Potentially Avoidable ER Visits	3M	
Domain 2, A	Potentially Avoidable Readmissions	3M	
Domain 2, A	PQI Suite – Composite of all measures	AHRQ	
Domain 2, A	PDI Suite – Composite of all measures	AHRQ	
Domain 2, A	CAHPS Measures (various)	AHRQ	
Domain 2, B	CAHPS Measures (care coordination with provider…)	AHRQ	
Domain 3, A (BH)	All Claims and MDS-based Metrics listed in Attachment J	3M, NCQA, CMS	Claims, Medical Rec, MDS
Domain 3, B (CVD)	All Claims Metrics listed in Attachment J	AHRQ, NCQA, CAHPS,	Claims, Survey, Medical Rec
Domain 3, C (Diabetes)	All Claims Metrics listed in Attachment J	AHRQ, NCQA, CAHPS	Claims, Medical Rec, Survey
Domain 3, D (Asthma)	All Claims Metrics listed in Attachment J	AHRQ, NCQA	Claims
Domain 4	Age-adjusted preventable hospitalizations rate per 10,000-Aged 18+ years		SPARCS
Domain 4	Asthma emergency department visit rate per 10,000		SPARCS
Domain 4	Asthma emergency department visit rate per 10,000 (aged 0-4)		SPARCS
Domain 4	Age-adjusted heart attack hospitalization rate per 10,000		SPARCS
Domain 4	Rate of hospitalizations for short-term complications of diabetes per 10,000 (aged 6-17 years)		SPARCS
Domain 4	Rate of hospitalizations for short-term complications of diabetes per 10,000 (aged 18+ years)		SPARCS

**Information in table is taken directly from Attachment J, when completed.*

d) ***Data Collection: This discussion shall include: A description of the data sources; the frequency and timing of data collection; and the method of data collection. The following shall be considered and included as appropriate:***
   - i. ***Medicaid encounter and claims data in TMSIS,***
   - ii. ***Enrollment data,***
   - iii. ***EHR data, where available***
   - iv. ***Semiannual financial and other reporting data***
   - v. ***Managed care contracting data***
   - vi. ***Consumer and provider surveys, and***
   - vii. ***Other data needed to support performance measurement***

Time Series Analysis

All datasets discussed in this proposal are available through the NYSDOH. The process of accessing the data (e.g., Medicaid claims, SPARCS, etc.) would begin immediately following the start date of this Agreement (if not sooner). Once obtained, data cleaning, management, and analyses would begin and continue throughout the duration of the evaluation. See Gantt Chart in the timeline below for further details.

Qualitative Analysis

Process/Implementation Study Data Collection
There will be several phases for identification of the sample. The first phase is identification of the provider and administrator sample. Using purposive sampling, the researchers will specifically use "critical or typical case sampling" (Bryman 2012; Creswell 2013; Patton, 2002). Each case, or PPS, is chosen because they demonstrate a specific pathway to DSRIP implementation and process (Ritchie, Lewis, 2014).

Identification of providers/administrators: There are 25 PPS located throughout four New York State regions (Metropolitan, Capital Area, Central, Western) (See Table 3). Each of the PPS has a primary, and often a secondary contact. Researchers will first identify the appropriate central contact and will work closely with the PPS staff and administrators to identify the appropriate stakeholders needed for interviews and focus groups. PPSs will aid the research team by providing lists of names and contact information for appropriate PPS planners and administrators for interviews from the providers within the PPS. In addition, lists of names and contact information (including e-mail addresses) will be sought from PPSs identifying relevant providers that are associated with and knowledgeable of each of their DSRIP projects. This information is necessary for the administration of surveys addressing specific projects. Because provider lists are so vast within the PPS, identifying the appropriate stakeholders is important as it will guide recruitment efforts for focus groups, with the goal of recruiting a diverse group of perspectives.

Identification of patients: The second phase is identification of patient sample. Purposive sampling will also be used but the approach to collection of the sample will be to develop a heterogeneous sample where the researches can include cases which vary widely while also determining themes that cut across the variety of cases (Ritchie, Lewis 2014).

Patients from each PPS who are part of DSRIP related services will be eligible for surveys. Each PPS will work collaboratively within its administration and with the researchers to identify a contact list of eligible patients. Inclusion criteria will include male and female patients' age 18 and up and those who

have not opted out of DSRIP-related data collection and protection. Although a web based survey will be utilized, internet access will not be inclusion criteria because researchers will mail surveys to the sample

Table 4: PPSs by region and location

DOH regions	PPS	Location
Capital Area	Adirondack Health Institute, Inc.	Queensbury
Capital Area	Albany Medical Center Hospital	Albany
Capital Area	Alliance for Better Health Care, LLC	Albany
Capital Area	Bassett Medical Center	Cooperstown
Central	Central New York Care Collaborative, Inc.	Syracuse
Central	Samaritan Medical Center	Watertown
Central/Western	Southern Tier Rural Integrated Performing Provider System, Inc.	(Broome, Cayuga, Chemung, Chenango, Cortland, Delaware, Schuyler, Steuben, Tioga and Tompkins
Western	Finger Lakes Performing Provider System, Inc.	Rochester
Western	Millennium Collaborative Care	Buffalo
Western	Sisters of Charity Hospital of Buffalo, New York	Buffalo
Metropolitan	Advocate Community Providers, Inc.	NYC
Metropolitan	Bronx-Lebanon Hospital Center	Bronx
Metropolitan	Maimonides Medical Center	Brooklyn
Metropolitan	Montefiore Medical Center	Bronx
Metropolitan	Mount Sinai PPS, LLC	NYC
Metropolitan	Nassau Queens Performing Provider System, LLC	Long Island
Metropolitan	New York City Health Hospitals Corporation	NYC
Metropolitan	NYU Lutheran Medical Center	Brooklyn
Metropolitan	Refuah Community Health Collaborative	Rockland/Orange Counties
Metropolitan	SBH Health System	Bronx
Metropolitan	Staten Island Performing Provider System, LLC	Staten Island
Metropolitan	State University of New York at Stony Brook University Hospital	Stony Brook
Metropolitan	The New York and Presbyterian Hospital	NYC
Metropolitan	The New York Hospital Medical Center of Queens	Queens
Metropolitan	Westchester Medical Center	Westchester

Sample Collection Frequency and Timing

Focus groups
Focus groups will be held once in each of the PPS groups per the entire evaluation period (Evaluation years 1-4).
Interviews
Key informant interviews will be conducted with PPS staff and administrators annually (Evaluation years 1-4).
Surveys for Providers
Surveys will be conducted with DSRIP-associated providers once per year (Evaluation years 1-5).
Surveys for Patients
Surveys will be conducted with patients once per year (Evaluation years 1-5).

e) Assurances Needed to Obtain Data: The design report will discuss the state's arrangements to assure needed data to support the evaluation design are available.

Addressing the evaluation questions will involve the use of a number of existing data sources that are maintained by, or are available to, the New York State Department of Health. Descriptions of each, including their availability, are as follows.

Medicaid Claims
This database contains billing records for health care services, including pharmacy, for approximately 5.7 million individuals enrolled in Medicaid in a given year. Also included are data on Medicaid enrollment status, diagnoses and provider associated with the billed services. The Medicaid claims database is updated on a monthly basis to include additional claims and modifications to existing claims. Given the claims processing, there is a 6-month lag in the availability of complete and finalized Medicaid claims data, where data for a given year are considered final by June 30th of the following year.

Medicare Claims
For the approximately 15% of Medicaid enrollees who are dually eligible for Medicare, Medicare claims will be used to ensure data completeness, as many of the services received by this group will be paid by Medicare and thus not appear in the Medicaid database. Medicare claims contains billing records for health care services, including pharmacy services, along with data on diagnoses and provider information. NYSDOH is working with an external entity specializing in the linking of Medicaid and Medicare claims data under a Coordination of Benefits Agreement (COBA), which will ensure timely access to Medicare claims through monthly data updates.

Statewide Planning and Research Cooperative System (SPARCS)
The Statewide Planning and Research Cooperative System (SPARCS) is an all payer data reporting system established in 1979 as a result of cooperation between the health care industry and government. Initially created to collect information on discharges from hospitals, SPARCS currently collects patient level detail on patient characteristics, diagnoses and treatments, services, and charges for inpatient and outpatient (ambulatory surgery, emergency department, and outpatient services), hospital services and outpatient services from free-standing ambulatory surgery centers. SPARCS data may be used for medical or scientific research or statistical or epidemiological purposes. All entities seeking SPARCS identifiable or limited data must submit a request to SPARCS Operations using standard data request forms. Finalized SPARCS data for a given year are available in August of the following year.

Minimum Data Set (MDS)

MDS 2.0 and 3.0 data consist of federally mandated assessments collected at regular intervals on all nursing home residents in New York. Assessment data collected include diseases and conditions, nutritional status, resident physical and cognitive functioning (e.g., activities of daily living), medications received, and nursing home admission source and discharge disposition. These data have been shown to be adequately reliable and are widely used in research, and are available to the New York Department of Health under data use agreement with CMS. There is, approximately, a 6-month lag in the availability of complete MDS data, where finalized data for a given year are available in June of the following year.

Consumer Assessment of Healthcare Providers and Systems (CAHPS®)

The Clinician & Group version of the CAHPS® survey will be administered by NYSDOH annually during the DSRIP demonstration period and will serve as the data source for selected outcome measures. The survey is administered by both mail and telephone, and assesses patients' experiences with health care providers and office staff. This includes information on patient experience over the last twelve months including most recent visit to provider, ease of getting an appointment, and wait times while in the office. The survey includes standardized questionnaires for adults and children. The adult questionnaire can be used in both primary care and specialty care settings; the child questionnaire is designed for primary care settings, but could be adapted for specialty care. Users can also add supplemental items to customize their questionnaires. Surveys are administered in September of a given year, and are available for use in February of the following year. Given confidentiality agreements, only de-identified CAHPS data will be available for use.

New York Vital Statistics

Birth and death certificate data are maintained by New York, with New York City Department of Health and Mental Hygiene and the New York Department of Health comprising two separate jurisdictions in the reporting of birth and death records, which will likely necessitate separate data use agreements. NYSDOH has the responsibility for annual statewide reporting of vital statistics governed by the terms of a memorandum of understanding between the two jurisdictions. Birth records contain information such as maternal medical risk factors, prenatal care received, infant birth date, birth weight, and infant diseases/conditions including congenital malformations. Death certificate data include date of death, underlying and multiple cause of death, decedent demographics, county of residence, and county of death. While Vital Statistics data are received by NYSDOH on an ongoing basis, due to the process of updating and finalizing information from birth and death certificates (e.g., due to delayed receipt of lab results), data for a given year are not considered complete until the end of the following year. In the event that the Independent Evaluator cannot be granted permission for vital statistics, either all or in part, NYSDOH will provide aggregated figures as needed for purposes of the DSRIP evaluation.

f) Data Analysis: This includes a detailed discussion of the method of data evaluation, including appropriate statistical methods that will allow for the effects of the Demonstration to be isolated from other initiatives occurring in the state. The level of analysis may be at the beneficiary, provider, health plan and program level, as appropriate, and shall include population and intervention-specific stratifications, for further depth and to glean potential non-equivalent effects on different sub-groups. Sensitivity analyses shall be used when appropriate. Qualitative analysis methods shall also be described, if applicable.

Process/Implementation Study Data Analysis

The qualitative data obtained through key informant interviews, focus groups, and open-ended survey questions will be transcribed and analyzed, using a qualitative data software program. Once data are organized and reviewed, researchers will use an integrated approach to identify and categorize the data according to concepts, relationships between concepts, and evaluative participant perspectives. Categorization based on setting and participant characteristics will also be completed, as appropriate. This categorization process facilitates the development of taxonomies, themes and theory, and comparisons. Responses will then be reviewed independently by at least two evaluation team members utilizing the finalized coding structure. Any coding discrepancies between reviewers will be resolved with discussion to achieve consensus. Coded data will be analyzed and interpreted to identify major concept domains and themes.

Quantitative survey data will be analyzed using SPSS statistical software.

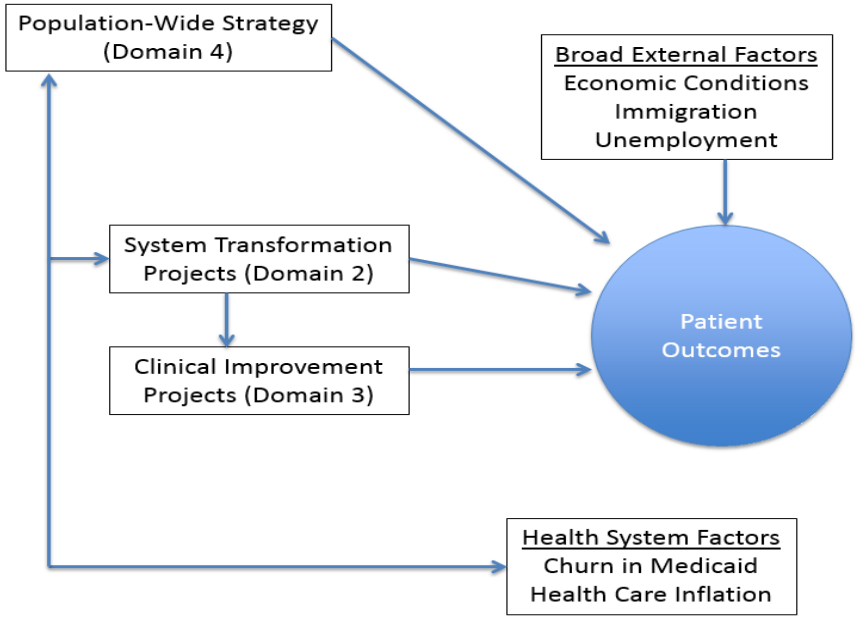
Analysis will focus on identifying usable feedback for improvement for each of the 25 PPSs. An additional focus will be identifying common and unique themes that arise in the data to inform the evaluation of DSRIP implementation as a whole.

Comparative Analysis

Conceptual Framework for Comparative Analysis.

Figure 1 presents a conceptual framework depicting the factors that are expected to impact health outcomes in the broader context of the DSRIP program. System Transformation (Domains 2), Clinical Improvement (Domain 3), and Population-wide Strategies (Domain 4) are all anticipated to impact patient-level outcomes. Moreover, broad external factors, such as economic conditions, immigration, and unemployment, are also likely to influence patient outcomes. To this point, issues related to beneficiary eligibility and the frequency of patients going in and out of the Medicaid system tend to play a role in influencing health outcomes. In addition, the varying performance levels and culture related to organizations that are early adopters versus late adopters of DSRIP projects and strategic initiatives also are likely to play a role in determining patient-level outcomes.

Figure 1. Conceptual Framework.

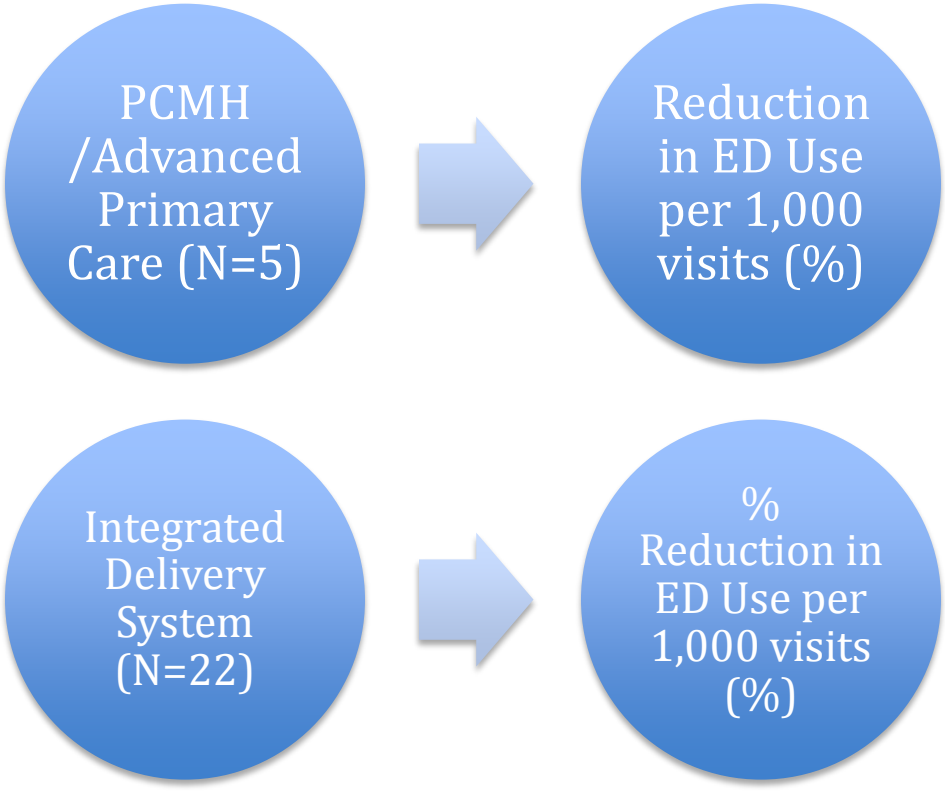


Examples of Multi-level Modeling to be used in Comparative Analysis

Evaluating the DSRIP, given the multiple PPS networks, partnerships, and projects within each domain is a complex endeavor. The evaluation team will leverage both qualitative and quantitative data to inform the evaluation design by embracing the variation across and within PPS interventions and the varied goals of each.

Early analyses will focus on the direct relationship between domain projects and the ultimate outcome measure. Analyses will be descriptive in nature when examining broader PPS outcomes, but additional multivariate analysis will be used to control for differences between populations, regions, providers, and other characteristics of the PPS that exist beyond the intervention or within the intervention project.

Figure 2: Descriptive Analysis Example for Domain 2 Impact on Emergency Department Visits.



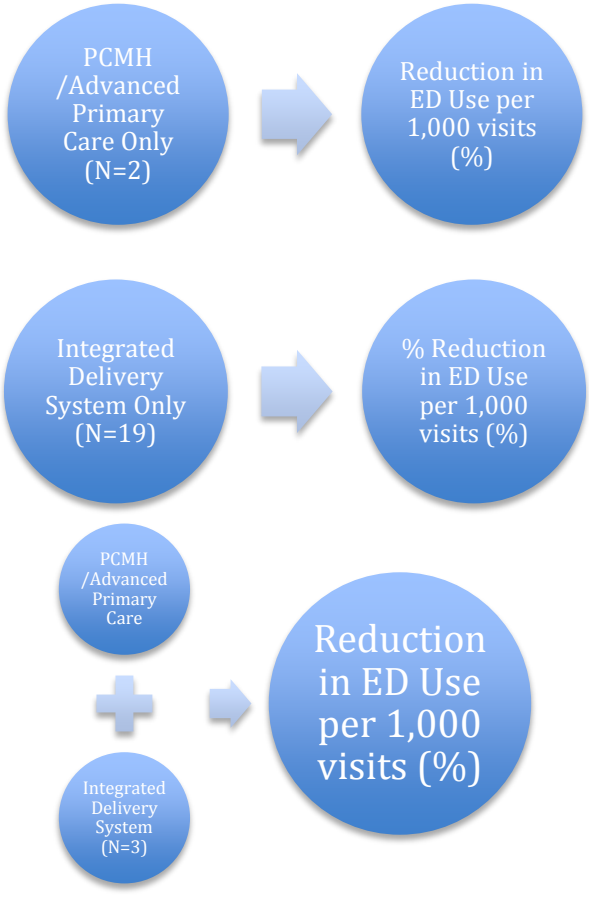
In the example presented in Figure 2, the underlying hypotheses are that specific Domain 2 projects will result in reductions in the percentage of Emergency Department (ED) visits per 1,000 total visits over time (from pre-DSRIP to post-DSRIP) in aggregate. Testing this hypothesis will simply use the inventory of DSRIP projects across PPS (see *Table Y: DSRIP Performing Provider System Project Selections*) and use descriptive statistics to understand if the percent change in ED visit use was reduced in the five PPS that had a PCMH/Advanced Primary Care intervention when compared to sites without a PCMH/Advanced Primary Care intervention, and separately calculate whether the 22 PPS with an Integrated Delivery System intervention experienced a reduction in ED visits when compared to those without an Integrated Delivery System intervention. These descriptive tables will give a general sense of what happened for the groups of sites that opted into a specific Domain project versus those that did not, but does not address multiple interventions in the same domain or control for underlying PPS characteristics. The unit of analysis will be the PPS site and data will be pulled from the PPS project list and administrative records (Medicaid claims for ED visits) and/or Quarterly Reports (from the PPS to NYSDOH). The resulting table is likely to appear in the evaluation report in the following format:

Table 5. Example Output for Bivariate Analysis by Project

Domain 2 Project	Number of Participants	Measure 1: Percentage Change in ED Visits Per 1,000					
		Baseline Rate	Year 1	Year 2	Year 3	Year 4	Year 5
1. Integrated Delivery System	22	1.3 per 1,000 visits	-0.2	-0.3	-0.4	-0.3	-0.5
2. PCMH/Advanced Primary Care	5	1.1 per 1.000 visits	-0.1	-0.2	-0.3	-0.2	-0.4

The second stage of descriptive analysis will focus on interactions between Domains and Projects between PPS networks, to better understand the impact of the customizability and flexible nature of the DSRIP interventions we are tasked with evaluating.

Figure 3. Descriptive Analysis for Multiple Domain 2 Projects' Impact on Emergency Department Visits



In the example in Figure 3, the additive relationship of implementing a PCMH/Advanced Primary Care project along with an Integrated Delivery System project can be better understood and incorporated into the evaluation approach. The resulting table is likely to appear in the evaluation report in the following format:

Table 6. Example Output for Bivariate Analysis by Project Combinations

Domain 2 Project	Number of Participants	Measure 1: Percentage Change in ED Visits Per 1,000					
		Pre-DSRIP Rate	Year 1	Year 2	Year 3	Year 4	Year 5
1. Integrated Delivery System Only	19	1.2 per 1,000 visits	-0.2	-0.3	-0.4	-0.3	-0.5
2. PCMH/Advanced Primary Care Only	2	1.0 per 1.000 visits	-0.1	-0.2	-0.3	-0.2	-0.4
2 & 3. PCMH/Advanced Primary Care + Integrated Delivery System	3	1.4 per 1,000	-0.2	-0.2	-0.3	-0.4	-0.5

In both of the examples above (Figures X.1 and X.2), the unit of analysis is the PPS, with the projects aligned with aggregate measures of ED visits reported or calculated at the PPS level. However, we also plan to leverage the individual level data when possible to understand the independent effects of each project on patient-level outcomes by controlling for individual patient characteristics for the beneficiaries nested within each PPS, and developing multivariate models to predict ED use over time using the Medicaid claims data to understand ED use for each individual. The regression analysis could focus on the rate of change in ED use over time, but because ED use is a fairly rare outcome at an individual level (more than half of subjects may have no ED use at all in a given year (Kaiser Family Foundation: http://kff.org/other/state-indicator/emergency-room-visits-by-ownership/), it would make more sense to use a two-step model predicting ED use (binomial logistic regression) and a conditional model (log-link Poisson or GLM model) for those with any ED use predicting the number of ED visits over time for each individual. Each individual would be nested in a PPS based on where they are attributed according to administrative records, and the qualitative data or progress reporting would be used to assign PPS values to capture categories of projects and/or variation in the interventions within project. While there are not sufficient degrees of freedom to do regression analysis at the PPS level, the individual level data would provide substantial data to test hypotheses about population health outcomes and measure change as a result of the DSRIP overall and individual projects or combinations of projects. The resulting regression equations would be based upon the distribution of the data and variables from the multiple data sources available to the evaluation team. The two-step model would be based upon the following general theory:

Step 1: Binomial Logistic Regression Predicting Any ED Use

$$Y_{ipt} = ß_0 + ß_{pt}D2PROJ1_t + ß_{pt}D2PROJ2_t + ß_iRACE_i + ß_{it}AGE_{it} + ß_iGENDER_i + ß_iILLNESS_i + ß_{it}AIDCODE_{it} + ß_{it}MONTHS_{it} + ε$$

where:
y = Presence of any Emergency Department visit during year
D2PROJ1 = Domain 2, Project 1 (Integrated Delivery System)
D2PROJ2 = Domain 2, Project 2 (PCMH/Advanced Primary Care)
ILLNESS = Presence of a chronic illness
AIDCODE = Medicaid aid code assigned by eligibility worker for a 12-month period
MONTHS = total number of months enrolled in Medicaid in a given year
i = individual

p = Performing Provider System Setting
t = year
ε = error term

Step 2: Log-Link Poisson Regression Predicting Number of ED Visits

$N_{ipt} = ß_0 + ß_{pt}D2PROJ1_t + ß_{pt}D2PROJ2_t + ß_iRACE_i + ß_{it}AGE_{it} + ß_iGENDER_i + ß_iILLNESS_i + ß_{it}AIDCODE_{it} + ß_{it}MONTHS_{it} + ε$

where:
N = Count of Emergency Department visits in year
D2PROJ1 = Domain 2, Project 1 (Integrated Delivery System)
D2PROJ2 = Domain 2, Project 2 (PCMH/Advanced Primary Care)
ILLNESS = Presence of a chronic illness
AIDCODE = Medicaid aid code assigned by eligibility worker for a 12-month period
MONTHS = total number of months enrolled in Medicaid in a given year
i = individual
p = Performing Provider System Setting
t = year
ε = error term

Clustering to create PPS comparison groups. Our approach will begin by clustering PPSs to compare those that have adopted specific domains and projects within those domains versus those that did not. More specifically, this will allow us to understand broadly, the impacts of PPSs that elected projects addressing asthma care to those that did not. A second approach we will use is to cluster PPSs based on their Domain 2 and Domain 3 selections. For example, several PPS's selected 2.b.iv. (Care Transitions to reduce 30-day readmissions) and 3.b.i (Evidence-based strategies for disease management in high-risk/affected populations), whereas others selected 1 of the above or neither. We would cluster these groups of PPS's to create comparison groups and examine specific metrics, such as readmission rates, etc. This approach will identify the potentially most impactful Domain 2 and 3 projects.

Tests of statistical significance will be used to determine whether material differences exist between PPSs. For measures available at the aggregate level for each PPS, we can only examine the bivariate association between the presence of a specific domain or project (or the level of implementation for that project) and the outcome variable. In that case, we will employ chi-square analysis to understand if differences are significant. However, in the case that outcome variables are available at the individual level (from Medicaid claims, for example), we can control for patient characteristics via multivariate, multilevel modeling because we will have individuals nested via attribution in each PPS.

Then, to provide further context for these findings, we will use the key informant interview and survey data previously gathered by CHSR to contextualize "how" certain PPSs have implemented project-specific plans and better understand "why" certain strategies may have been more or less effective in the context of comparative analysis.

Difference-in-differences. We will use a difference-in-differences estimation methodology to examine specific performance measures in the time before and after the implementation of the DSRIP program comparing PPSs involved in specific interventions to those that were not engaged in those projects. This estimation strategy adjusts for time-based variations in outcomes, helping determine program impacts from other phenomena. Moreover, this approach will give us an aggregate understanding as to whether

the overall picture has changed for specific domains based on key measures of interest defined in Attachment J (http://www.health.ny.gov/health_care/medicaid/redesign/dsrip/cms_official_docs.htm).

This approach will also require the use of risk-adjusted measures. This will be ideal because it would level the playing field in terms of the dual-eligibles & SSI patients as these individuals tend to seek care at distinct locations and are typically-high utilizers of care. Also, prior to carrying out this analysis, we will endeavor to identify patients and providers (hospitals and medical groups) who were not involved in any DSRIP PPS and understand the trends in use, quality, and spending over time in a separate difference-in-differences analysis.

Patient-level Comparisons. We will examine trends within and across PPSs with respect to patient-level outcomes. In particular, we will focus such comparisons on factors including age, sex, race, presence of chronic conditions, and mental health/substance abuse to inform our understanding of patients' service experience and satisfaction during the DSRIP program. Such analyses will require the use of CAHPS data to examine patient satisfaction scores. However, because CAHPS scores/responses are typically not attributed to specific patients and are only available at the department, hospital, medical group, physician, or health plan level, we will need to examine the organizational-level CAHPS scores and their relationship to patient-level outcomes for populations attributed to the specific organization (at multiple levels). To effectively conduct such an analysis, we will build upon the approach set forth by Sequist et al. (2008) to deal with the lack of individual-level outcome data linked to CAHPS scores.

Because we know the Medicaid population can be vulnerable to income status changes and other reasons for disenrollment, we will determine inclusion criteria based upon months enrolled over each 12 month time period for specific measures (for example, HEDIS-based quality measures often require 11 months of enrollment) and gaps in coverage. When considering other measures, like spending and patient experience, all Medicaid members will be included for the months they were enrolled over the 36 month program and the 12 month look-back period for pre-DSRIP data.

Data. Given our interest in the variables in Table 2, our research team has identified the following data sets that will aid in our comparative analyses. They are as follows:

1. **Medicaid & Medicare Claims.** Medicaid and Medicare claims data will be the primary source of data for our analyses. This data will houses the details related to many of the metrics referenced in Table D.3.
2. **SPARCS.** The data related to a number of the aforementioned measures is stored in the SPARCS database. Use of this data will allow us to investigate key metrics and compare across PPSs.
3. **MDS (long-term care).** For measures specific to long-term care (e.g., Domain 3, Behavioral Health, % of Long Stay Residents who have Depressive Symptoms).
4. **CAHPS.** The use of CAHPS data will allow us to learn about variations in service experience and patient satisfaction during the DSRIP program and examine the linkage between organization-level patient experience and individual-level outcomes.
5. **Vital Statistics.** Birth records and death certificate data will also be used in our analyses and include variables such as prenatal care received, date of death, county of residence, etc.

g) Timeline: This includes a timeline for evaluation-related milestones, including those related to procurement of an outside contractor, if applicable, and deliverables.

Chart 1. Timeline of Procurement Process, Research Activities & Evaluation Reporting Requirements

NYS DSRIP Program Evaluation Timeline		2015	2016				2017				2018				2019				2020				2021
		Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1
		DY1		DY2				DY3				DY4				DY5							
Procurement & Research Activities	Request for proposals released																						
	Application received																						
	Applications reviewed, winning bidder identified																						
	NYS contract approval process																						
	Contract finalized																						
	Develop/design protocols for IRB submission																						
	IRB submission																						
	DUA for Medicaid and other data executed																						
	Schedule & perform key informant interviews																						
	Schedule & perform focus groups																						
	Transcribe, code, and analyze interview and focus group text																						
	Design web-based survey																						
	Administer web-based survey																						
	Analyze web-based survey data																						
	Receive Medicaid claims data																						
	Submit request for SPARCS and other data																						
	Receive SPARCS and other data																						
	Data cleaning and preparation																						
	Data analysis																						
Reporting Requirements	Quarterly Update Reports to NYSDOH																						
	Annual PPS Reports																						
	Annual Statewide Reports																						
	Draft Interim Evaluation Report Due to NYSDOH																						
	Draft Interim Evaluation Report Due to CMS																						
	Final Interim Evaluation Report Due to NYSDOH																						
	Final Interim Evaluation Report Due to CMS																						
	Preliminary Summative Evaluation Report Due to NYSDOH																						
	Preliminary Summative Evaluation Report Due to CMS																						
	Final Draft Summative Evaluation Report Due to NYSDOH																						
	Final Draft Summative Evaluation Report Due to CMS																						
	Final Summative Evaluation Report Due to NYSDOH																						
	Final Summative Evaluation Report Due to CMS																						

h) Evaluator: This includes a discussion of the state's process for obtaining an independent entity to conduct the evaluation, including a description of the qualifications that the selected entity must possess; how the state will assure no conflict of interest, and a budget for evaluation activities.

The procurement process to contract with an independent entity to conduct the evaluation began in December 2015. In a competitive bidding process, a Request for Proposals (RFP) was developed and issued by the New York State Department of Health. This RFP described the scope of work, the major tasks, and contract deliverables, with a period during which potential applicants could submit questions to the Department. Proposals received were reviewed by a panel of NYSDOH staff, using a scoring system developed for this RFP. Eligible bidders could not be employees or entities of the New York Department of Health, and not have any business relationship with any of the PPS's or their participating providers. Applicants were evaluated on the basis of related work experience, staffing level and expertise, environment and resources, data analytic capacity, and ability to act as an independent, unbiased third party in conducting the evaluation. The contract with the winning bidder was finalized in December 2016.