This report is part of the RAND Corporation research report series. RAND reports present research findings and objective analysis that address the challenges facing the public and private sectors. All RAND reports undergo rigorous peer review to ensure high standards for research quality and objectivity.

Project Report

CMS Innovation Center Health Care Innovation Awards

Evaluation Plan

Sandra H. Berry, Thomas W. Concannon, Kristy Gonzalez Morganti, David I. Auerbach, Megan K. Beckett, Peggy Guey-Chi Chen, Donna O. Farley, Bing Han, Katherine M. Harris, Spencer S. Jones, Hangsheng Liu, Susan L. Lovejoy, Terry Marsh, Grant Martsolf, Christopher Nelson, Edward N. Okeke, Marjorie L. Pearson, Francesca Pillemer, Melony E. Sorbero, Vivian L. Towe, Robin M. Weinick

Sponsored by the Centers for Medicare and Medicaid Services

The research described in this report was sponsored by the Centers for Medicare and Medicaid Services, and was produced within RAND Health, a division of the RAND Corporation.

RAND OFFICES

SANTA MONICA, CA • WASHINGTON, DC

PITTSBURGH, PA • NEW ORLEANS, LA • JACKSON, MS • BOSTON, MA

DOHA, QA • CAMBRIDGE, UK • BRUSSELS, BE

Preface

The Health Care Innovation Awards (HCIAs), sponsored by the Center for Medicare and Medicaid Innovation (CMMI) within the Centers for Medicare & Medicaid Services (CMS), are intended to support the development of compelling new models of service delivery or payment improvements that promise to deliver better health, better health care, and lower costs through improved quality of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) enrollees. In May and June 2012, CMS announced the recipients of the HCIAs, with awards to 107 awardees ranging from approximately $1 million to $30 million.

CMS contracted with the RAND Corporation to carry out the development of an overall evaluation strategy for the HCIA awardees. This report describes that strategy. A related report, *The CMS Innovation Center Health Care Innovation Awards Database Report: Information on Awardees and their Populations* (Morganti et al., 2013), provides the results of RAND's effort to collect information on each of the awardees. This report is written for use by CMS staff who will be engaged in planning the evaluations, by individuals and organizations that will conduct or support evaluation activities, and by awardees who will participate in evaluations by sharing information on HCIA programs and program outcomes.

Timothy Day serves as the CMS-CMMI contracting officer's representative. This research was conducted by RAND Health, a division of the RAND Corporation. A profile of RAND Health, abstracts of its publications, and ordering information can be found at http://www.rand.org/health.

Table of Contents

Preface ... iii
Figures ... vi
Tables ... vii
Summary ... ix
- Goal and Challenges for Evaluation ... ix
- An Evaluation Strategy for HCIA Awardees ... x
  - Evaluation at the Level of the Individual Awardee ... x
  - Evaluation Within Awardee Groupings ... xi
  - Evaluation Across All Awardees and Groupings ... xii
- Evaluation Dimensions, Measures, and Designs ... xiii
  - Conceptual Framework ... xiii
  - Key Dimensions Within the Framework ... xiv
- Summary Evaluation and Decision Strategy ... xviii
  - Strategy ... xix
- Conclusion ... xx

Acknowledgments ... xxi
Abbreviations ... xxii
Chapter 1. Background, Evaluation Goals, and Overview ... 1
- Background: Policy Context for the Health Care Innovation Awards ... 1
- An Evaluation Strategy for HCIA Awardees ... 2
  - Goal of the Evaluation ... 2
  - Challenges in Designing an Evaluation Strategy ... 2
- Overview of the Document ... 3

Chapter 2. Strategies for Evaluation at Three Levels ... 5
- Evaluation of Individual Awardees ... 5
- Evaluation Within Awardee Groups ... 6
  - Management of Medically Fragile Patients in the Community ... 8
  - Hospital Setting Interventions ... 8
  - Community Interventions ... 8
- Summary Evaluation ... 9

Chapter 3. Evaluation Dimensions, Measures, and Designs ... 11
- Conceptual Framework ... 11
  - Key Dimensions Within the Framework ... 12
- Implementation Effectiveness ... 14
  - What Is Implementation Effectiveness? ... 14
  - Why Is It Important to Measure Implementation Effectiveness? ... 14
  - How Can We Measure Implementation Effectiveness? ... 15

Program Effectiveness 19
What Is Program Effectiveness? 19
Why Is It Important to Measure Program Effectiveness? 20
How Can We Measure Program Effectiveness? 21
Workforce Issues 42
What Are Workforce Issues? 42
Why Are Workforce Issues Important for the HCIA Project? 43
How Might Workforce Issues Be Measured? 43
Impact on Priority Populations 47
What Are Priority Populations? 47
Why Is It Important to Evaluate Priority Populations? 48
How Might We Measure Priority Populations? 48
Context 51
What Is Context? 51
Why Is It Important to Measure Context for the HCIA Project? 51
How Might Context Be Measured? 52
Chapter 4: Summary Evaluation and Decision Strategy 55
Challenges 55
Strategy 63
Coordination of Evaluators 63
Analysis and Interpretation of Findings 64
Conceptual Framework and Analysis and Interpretation of Findings 65
Component 1: A Ratings System 66
Evaluation Dimensions 67
Development of a Common Scale and Ratings System 72
Development of Uniform Strength of Evidence Indicators 73
Component 2: A Pooled Analysis 75
Advantages 75
Data Sources and Measures 85
Comparison Groups 85
Interpretation of Impact Estimates 86
Component 3: A Decision Strategy 86
The Decision Process 88
Conclusions 91
Appendix A. Relationship of Evaluation Dimensions to HCIA Evaluation Questions 92
Appendix B. Sample Size Requirements for Estimating HCIA Impacts 95
Medical Care Utilization 96
Medical Care Costs 99
Other Self-Reported Outcomes 103
References 106

Figures

Figure S.1. Conceptual Framework xiv
Figure 3.1. Conceptual Framework 12
Figure 4.1. Expanded Conceptual Framework 66
Figure 4.2. Projected Number of Participants by Award and Sample Size Requirements 77

Tables

Table S.1. Summary of Awardees Categories and Groupings xii
Table S.2. Evaluation Dimensions xv
Table 2.1. Summary of Awardees Categories and Groups 9
Table 3.1. Evaluation Dimensions 13
Table 3.2. Measurement of Program Drivers 15
Table 3.3. Measurement of Intervention Components, Dosage, Fidelity, and Self-Monitoring 17
Table 3.4. Measurement of Intervention Reach 19
Table 3.5. Measurement of Health 23
Table 3.6. Measurement of Cost 26
Table 3.7.1. Measurement of Safety 29
Table 3.7.2. Measurement of Clinical Effectiveness 30
Table 3.7.3. Measurement of Patient Experience 35
Table 3.7.4. Measurement of Timeliness 36
Table 3.7.5. Measurement of Efficiency 37
Table 3.7.6. Measurement of Care Coordination 38
Table 3.8. Measurement of Cross-Cutting Considerations 39
Table 3.9. Measurement of Workforce Issues 45
Table 3.10. Measurement of Priority Populations 49
Table 3.11. Measurement of Impact on Priority Groups 50
Table 3.12. Measurement of Context 52
Table 4.1. List of Awardee Interventions 56
Table 4.2. Implications of Savings and Sustainability for Policy and Practice 69
Table 4.3. Scoring Criteria 72
Table 4.4. The Ratings System 73
Table 4.5. Participants by Awardee 78
Table 4.6. The Ratings System 89
Table A.1. Crosswalk of HCIA Research Questions with Proposed Evaluation Dimensions 92
Table B.1. Evaluation Design Descriptions 95
Table B.2. Comparison with a Known, Fixed Benchmark (1-sample comparison) 96
Table B.3. Comparison Within the Same Group Pre and Post Intervention (paired comparison) 97
Table B.4. Cross-Sectional Comparison with a Control Group (2-sample) 98
Table B.5. Comparison with a Control Group (difference-in-difference) 99
Table B.6. Comparison with a Fixed Benchmark (1-sample comparison) 100

Table B.7. Comparison Within the Same Group Pre- and Post-Intervention (paired comparison) 101
Table B.8. Cross-Sectional Comparison with a Control Group (2-sample) 102
Table B.9. Comparison with a Control Group (difference-in-difference) 103
Table B.10. Comparison with a Fixed Benchmark (1-sample comparison) 104
Table B.11. Comparison Within the Same Group Pre- and Post-Intervention (paired comparison) 104
Table B.12. Cross-Sectional Comparison with a Control Group (2-sample) 105
Table B.13. Comparison with a Control Group (difference-in-difference) 105

Summary

On November 14, 2011, the Center for Medicare and Medicaid Innovation (CMMI) within the Centers for Medicare & Medicaid Services (CMS) announced the Health Care Innovation Challenge. Through this initiative, CMS planned to award up to $900 million in Health Care Innovation Awards (HCIAs), funded through the Affordable Care Act (ACA), for applicants who proposed compelling new models of service delivery or payment improvements that promise to deliver better health, better health care, and lower costs through improved quality of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) enrollees. CMS was also interested in learning how new models would affect subpopulations of beneficiaries (e.g., those eligible for Medicare and Medicaid and complex patients) who have unique characteristics or health care needs that could be related to poor outcomes. In addition, the initiative sought to identify new models of workforce development and deployment, as well as models that can be rapidly deployed and have the promise of sustainability.

This report describes a strategy for evaluating the awardees. It is written for use by the CMS staff who will be engaged in planning the evaluations, by individuals and organizations that will conduct or support evaluation activities, and by awardees who will participate in evaluations by sharing information on HCIA programs and program outcomes. A companion report, *The CMS Innovation Center Health Care Innovation Awards Database Report: Information on Awardees and their Populations* (Morganti et al., 2013), presents detailed information on each of the awardees.

Goal and Challenges for Evaluation

The goal of the evaluation is to help CMS answer two key questions:

- Which kinds of innovative approaches result in reduced cost while improving or maintaining the standard of care, patient health and quality of life, and satisfaction of the workforce?
- To the extent that a particular approach is promising, what contextual factors need to be in place to make success likely, and what contextual factors might cause serious problems?

There are complex challenges to designing an effective and comprehensive evaluation of the HCIA initiative. Below, we summarize a few of these challenges for evaluation design and implementation. All of these challenges will be addressed in the proposed strategy:

- **Evaluation skills of individual awardees.** Based on past experience with multiproject evaluations, we expect that the awardees funded under HCIA will

vary widely in terms of their capacity to provide information for carrying out the external evaluation. In order to conduct the evaluation, the external evaluator(s) will need to assess awardees' skill levels and establish a priority list for technical assistance.

- **Calculating program costs and cost savings.** A primary focus of program evaluators will be to identify models with the potential to reduce the total cost of care in ways that are financially sustainable for provider organizations. Key to achieving this aim is the ability of evaluators to measure effects of a comprehensive array of innovations on cost and quality outcomes. The ability to calculate program costs and cost savings will be challenging for many awardees.
- **Awardee time burdens and coordination.** Both in this evaluation design contract and the subsequent contract to perform the evaluation, it is essential to balance the need for valid evaluation data against the desire to minimize time and resource burdens on individual awardees. Our experience has demonstrated that it is preferable (whenever possible) for the evaluator to conduct secondary analyses of existing data and to supplement these analyses with awardee interviews or additional materials directly from their projects. This approach enables the evaluator to obtain complete data and eases the evaluation burden on awardees.
- **Cross-cutting versus awardee-specific measures**. A crucial issue for the evaluation strategy is to achieve a balance between the use of measures that cut across awardees' efforts and those specific measures relevant to each innovation model being tested. The correct balance will ensure that measures are general enough to inform cross-awardee comparisons while being relevant to a diverse array of awardee efforts and will also ensure that awardees do not bear an undue measurement burden.

An Evaluation Strategy for HCIA Awardees

The goal for the evaluation design process is to create standardized approaches for answering key questions that can be customized to similar groups of awardees and that allow for rapid and comparable assessment across awardees. The evaluation plan envisions that data collection and analysis will be carried out on three levels: at the level of the individual awardee, at the level of the awardee grouping, and as a summary evaluation that includes all awardees. The ultimate goal is to identify strategies that can be employed widely to lower cost while improving care.

Evaluation at the Level of the Individual Awardee

The first step in conducting an evaluation for each awardee will be to develop data at the level of the individual awardee. This may involve collection of program documents and other

materials, clinical data, self-report data from program patients or staff, or observational data (e.g., observations of key program activities being implemented).

Evaluation Within Awardee Groupings

In order to conduct evaluations at an operationally manageable level and to allow potential pooling of data for statistical analysis, RAND developed and CMS reviewed and approved groupings of awardees. We proposed a way of grouping awardees based on the larger questions the evaluation needs to answer, as well as on the day-to-day realities of how and in what parts of the care system the awardees are implementing their projects (i.e., approach and setting). We suggested grouping awardee projects across three larger categories:

- **Management of medically fragile patients in the community:** characterized by a focus on high-risk patients who are receiving care primarily in community health care settings with a goal of controlling costs by improving care quality and reducing emergency department (ED) visits and hospital admissions
- **Hospital setting interventions:** characterized by a focus on care of hospitalized patients with a goal of reducing the length of stay, intensity of utilization, duplication of care, and readmission
- **Community interventions:** characterized by a focus on care of beneficiaries in community settings, but focusing on various aspects of how care is delivered, rather than certain categories of patients—although some may also focus on subgroups of patients

While these three types of approaches are designed to improve quality of care and reduce or slow the growth of cost through better care, they will do so in different ways and with different specific end points, and these differences will need to be taken into account in designing an evaluation plan. It will also be important to capture the specific structural features of programs (e.g., health information technology [HIT] improvements, workforce training, payment reform); the processes they include (e.g., care coordination, patient navigation, home visitation, care standardization); the effects on specific clinical outcomes and health-related quality of life; and the specific ways in which they are affecting cost in terms of reduced intensity of care, reduced ED visits, reduced hospitalizations and readmissions, and other factors.

We proposed ten groupings for awardees within these three categories, as shown in Table S.1. Following discussions with CMS about the proposed groups and the assignment of awardees to the groups, RAND worked with CMS to finalize the assignment of awardees to the ten groupings.

Table S.1. Summary of Awardees Categories and Groupings

Category	Groupings
Management of medically fragile patients in the community	• Disease/condition-specific targeting (e.g., cardiac, asthma, dementia, diabetes, stroke, cancer, chronic pain, renal/dialysis) • Complex/high-risk patient targeting (e.g., multiple conditions, rural, low income, advanced illness) • Behavioral health patients being treated in community care settings
Hospital settings interventions	• Condition-specific targeting (e.g., sepsis, delirium) • Acute care management • Improvement in ICU care, remote ICU monitoring
Community interventions	• Community resource planning, prevention and monitoring • Primary care redesign • Pharmacy/medication management • Shared decisionmaking

NOTE: ICU = intensive care unit.

In addition to the grouping structure for the awardees, there are other characteristics that will be considered in the evaluation design recommendations and for the actual evaluation. These include

- target population characteristics (e.g., age, Medicare status, Medicaid status, and CHIP status)
- geographic region of the country and whether the area is urban or rural
- program size in terms of funding and number of beneficiaries
- workforce characteristics (e.g., type of staff, kinds of training and organization, and how types and levels are staff are deployed to undertake tasks within the system of care).

Evaluation Across All Awardees and Groupings

The value of a summary evaluation is the opportunity for CMS to examine aspects of program implementation, workforce, and context that may influence an intervention's effectiveness. We present several approaches for a summary evaluation of awardees and groupings. These include a meta-analytic approach, pooled data analyses, and a systematic ratings system. These approaches will help to identify intervention strategies that are most effective in reducing costs while improving quality of care. Finally, we present structured approaches for establishing consensus interpretations of awardee and grouping evaluations, as well as for arriving at decisions about which approaches are worth scaling up, which are worth studying further and which should be deferred from current consideration for further investment.

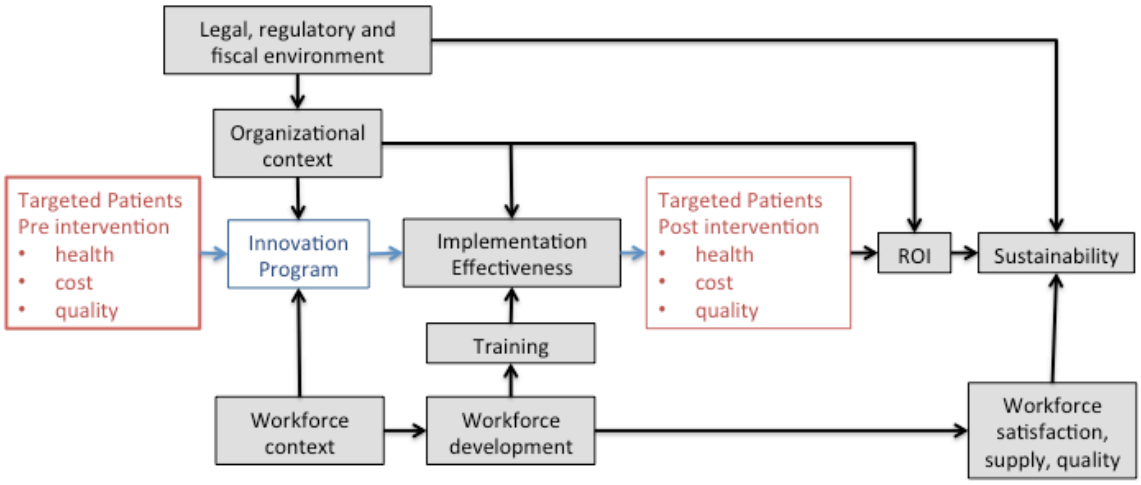
Evaluation Dimensions, Measures, and Designs

Conceptual Framework

The conceptual framework for the evaluation is shown in Figure S.1. The framework illustrates how key dimensions of the evaluation relate to a primary outcome of interest: the sustainability of an awardee program.

In the leftmost box, we depict the *health status and characteristics* of the target patient population. These characteristics motivate the *design of an innovation program*, which is also influenced by the *legal, regulatory, and fiscal environment*; the *organizational context*; and the *workforce context*. The *implementation effectiveness* of each program is affected by *organizational context* and *workforce training* and can be measured along four dimensions: *program drivers* (i.e., the theory behind the program and intended drivers of change); *intervention components* (e.g., training, technical assistance), *dosage* (i.e., the "amount" of the intervention delivered to patients or the health system), and *fidelity* (i.e., adherence to planned procedures); and the reach of the program. *Program effectiveness* is characterized by the evaluation dimensions of *health*, *cost*, and *quality*. All of these factors affect the *return on investment* (ROI), which, along with *workforce satisfaction*, affects the overall *sustainability* of the program. Each dimension in this framework represents a more complex set of elements. This framework is meant to be flexible so that it can be operationalized and interpreted by stakeholders with varying perspectives, including providers, evaluators, and CMS.

Figure S.1. Conceptual Framework



Key Dimensions Within the Framework

In Table S.2, we outline the key dimensions for the proposed evaluations. Below the table, we briefly define each of the dimensions and its importance for the HCIA project and explain the focus of measurement for the dimension.

Table S.2. Evaluation Dimensions

Category	Dimensions	Subdimensions
I. Implementation Effectiveness		
	A. Program drivers	1. Theory of change
		2. Theory of action
	B. Intervention	1. Components of the intervention
		2. Dosage
		3. Fidelity
		4. Self-monitoring
	C. Reach	1. Coverage
		2. Timeliness of implementation
		3. Secondary use of tools
II. Program Effectiveness		
	A. Health	1. Health outcomes
		2. HRQoL
	B. Costs	1. Program costs
		2. Utilization
		3. Expenditure
	C. Quality	1. Safety
		2. Clinical effectiveness
		3. Patient experience
		4. Timeliness
		5. Efficiency
		6. Care coordination
	D. Cross-cutting considerations	1. Equity and disparities
		2. Subgroup effects
		3. Spillover effects
III. Workforce Issues		
	A. Development and training	
	B. Deployment	
	C. Satisfaction	
IV. Impact on Priority Populations		
	A. Populations	1. Medical priority groups
		2. Nonmedical priority groups
	B. Impact	1. Cost reductions and savings
		2. Clinical outcomes
V. Context		
	A. Endogenous factors	1. Leadership
		2. Team characteristics
		3. Organizational characteristics
		4. Stakeholder engagement
	B Exogenous factors	1. Policy and political environment

Implementation Effectiveness

Implementation effectiveness refers to the degree to which an intervention is deployed successfully in real-world settings. Speed to implementation was a key consideration in the selection of HCIA awardees, and a key goal of the HCIA program is to identify innovations that can be rapidly deployed more widely once they have been determined to be effective.

Implementation effectiveness can be measured in terms of program drivers; intervention components, dosage, fidelity, and self-monitoring; and reach. *Program drivers* include the theory of change (i.e., the mechanisms that catalyze or otherwise cause changes in individual and organizational behavior) and the theory of action behind the intervention (i.e., the specific activities used to deliver the innovation). *Intervention components* include the specific activities by which the program seeks to induce better health outcomes at lower cost (e.g., training programs, patient navigators, HIT, new staffing). *Dosage* refers to how much of the innovation a health system or patient gets. *Fidelity* refers to how faithfully the innovation or program was delivered. *Self-monitoring* refers to awardee efforts to collect data on their own program activities and outcomes and the use of these data for quality improvement. *Reach* can be measured through the extent of the intervention's coverage (i.e., geographic reach, target population, number of individuals, organizations, or other units covered), the timeliness of its implementation, and the secondary use of tools that it generates.

Program Effectiveness

Program effectiveness refers to assessments of an intervention's impact on outcomes of interest, referring to the goals of reducing cost through better care and better health. HCIA awardees are expected to assess cost savings and to document improvements in health outcomes and quality over the three-year term of the award. They are also asked to project the intervention's effectiveness on an annualized basis after the term is finished.

We present three outcome dimensions that are of interest in health care innovation: health, costs, and quality. The health dimension focuses on the impact of the intervention on health outcomes, including mortality, morbidity, and health-related quality of life (HRQoL). The costs dimension focuses on program costs, impact on utilization, and changes in expenditures resulting from the intervention. The quality dimension focuses on improvements in care along several domains of quality: (1) safety, (2) clinical effectiveness, (3) patient experience, (4) timeliness, (5) efficiency, and (6) care coordination. We also discuss considerations that cut across the other dimensions in this section—including equity and health care disparities issues, effects on specific subgroups of interest, and spillover effects.

Workforce Issues

A critical challenge of delivery system reform is to identify and test new ways to create and support the workforce of the future—a workforce that will deliver and support new care models.

There are three key types of workforce issues to be considered: development and training, deployment, and satisfaction. In terms of development and training, it is important to understand what works best for implementation of the innovation: a training process and other strategies to add new skills to current workers or contracts with outside providers who already have those skills. How workers are deployed and how they interact with patients is also critical to the success or effectiveness of many of the awardees' interventions. Job satisfaction is key to providers' willingness to be part of this workforce, their ability to perform their work effectively, and the smooth functioning of a provider organization.

Key elements of development and training to be measured include the extent to which programs provide training to use existing staff and incorporate new kinds of staff effectively, the level of investment in training required to fill workforce gaps, and the effectiveness and efficiency of various training models. Deployment issues include the extent to which newly formed teams function together and the ways in which workforces are utilized in the innovation. To understand staff satisfaction, it is important to measure the extent to which different kinds and levels of staff are satisfied or dissatisfied with the care they are able to provide and with working conditions in general.

Impact on Priority Populations

Priority populations may include those with certain medical conditions, such as the chronically ill, pregnant women, persons with behavioral health needs, individuals with disabilities, and people living with HIV. Nonmedical priority populations might include senior citizens, children, low-income families, homeless individuals, immigrants and refugees, rural populations, ethnic/racial minority populations, non–English-speaking individuals, and underserved groups. Evaluating the impact of HCIA interventions on priority populations means understanding the potential impact of the intervention on these populations, including the impact on clinical outcomes and cost.

Two aspects of measuring intervention impact for priority groups are important: (1) the extent to which health outcomes, quality, and costs are different for individual priority groups compared to the health outcomes quality and costs for the intervention population as a whole and (2) whether outcomes, quality, and cost savings would be different for priority groups if the intervention were brought to full scale.

A number of metrics might be used to measure outcomes for priority groups. These include patient characteristics, mortality, morbidity, functional health status, HRQoL, technical quality, rating of providers, rating of provider communication, access to care, care coordination, courtesy and helpfulness of providers, cultural competency, self-management education, and rating of experience with new technologies and processes. In addition, it will be crucial to understand how cost impacts and population size may interact to produce potential savings.

Context

Context refers to the environment in which an innovation occurs and, more specifically, to the factors that can help facilitate or impede an intervention's success. Context includes such endogenous factors as leadership, team functioning, and organizational features and such exogenous factors as the policy and political environment in which an intervention is implemented. Key questions focus on the contextual factors that are needed to support a particular intervention: Were there unique characteristics of the awardee organization, market, approaches, or patient populations that affected the implementation and success of the innovation? Was there a clearly designated champion or leader to oversee implementation?

Key dimensions of context to be measured include endogenous factors (i.e., awardee characteristics, programmatic changes, leadership, team science, organizational issues) and exogenous factors, such as the policy and political environment. The relevant aspects of context will vary across interventions. Because they vary, we propose to assess context in terms of "fit" or "congruence" between two key elements: the demands and requirements of the innovation and the operational realities of the use context.

Summary Evaluation and Decision Strategy

In addition to the evaluations of individual awardees and awardee groups, we also see a role for summary evaluation strategies that would include other awardee groupings. For instance, a summary evaluation might assess awardees that include Medicare recipients as their primary target group. The primary objective of the summary evaluation is to compare and synthesize findings from evaluations conducted at the awardee and group levels, as well as from pooled analyses. The evaluations will assist in identifying (1) those interventions that can be implemented more broadly, (2) those that need testing in other settings, and (3) those that may be deferred from current consideration for further investment.

The benefits of a summary evaluation have to do with the potential to compare, synthesize, and interpret the variety of evaluations that are conducted on individual innovations and smaller groups of awardees. Comparison and synthesis can provide further insight on innovations that are effective at controlling or reducing costs and those that are effective at maintaining or improving health outcomes and quality of care. A summary evaluation can also provide data on how effective innovations can be scaled up to other populations and under what circumstances; what changes in regulations, reimbursement structure, and other policies may be needed to ensure the sustainability of effective innovations; and how less-effective innovations can be tested further, why their outcomes are lacking, and how their outcomes might be improved.

There are also several challenges associated with conducting a summary evaluation. The first of these has to do with the heterogeneity of awardee activities. Each awardee has proposed and is carrying out multiple, overlapping changes in its health care systems. Second, the awardees target a wide range of populations, and thus care must be exercised in interpreting the potential

for scale-up of successful innovations. Third, awardee innovations and their population impacts will be evaluated in the context of different organizational characteristics (e.g., differences in leadership support, information technology [IT], culture, staffing structure), which may be influential on outcomes. Fourth, and perhaps most challenging, individual awardees and evaluators may measure performance in different ways, which means that comparison and synthesis of measurement will be extremely challenging.

Strategy

The summary evaluation strategy has to take account of these challenges. Below we suggest key elements of a strategy that will create opportunities for valid comparison and synthesis of individual awardee and group evaluations.

Coordination of Evaluators

Early coordination of evaluators will be important because it can maximize correspondence and minimize unnecessary variation in the ways that awardee innovations have been assessed, through differences in evaluation questions, metrics, data, or approach. As awardee and group evaluations proceed, coordination will ensure that questions, metrics, data, and approaches are similar enough to produce findings that can be compared and synthesized across the many awardees, awardee group, and interventions. Coordination would begin with consideration of proposed evaluation dimensions. The process would continue with a discussion of the research questions, metrics, data, and approaches for evaluation within each of the awardee groupings.

Analysis and Interpretation of Findings

The analysis and interpretation approach we propose is composed of three major components, which can be carried out simultaneously.

Component 1: A Ratings System. An evaluation ratings system may be developed to compare findings from the many qualitative and quantitative measures in grouping, intervention, and program evaluations. This system could be focused on the five major evaluation dimensions presented earlier: implementation effectiveness, program effectiveness, workforce issues, impact on priority populations, and context. The characteristics are designed to summarize findings across evaluation dimensions, using different types of data.

Component 2: A Pooled Analysis. Further assessment of the interventions undertaken by awardees can be obtained via a pooled analysis using data from CMS, states, or other administrative or survey sources. The power of a pooled analysis is to combine observations from multiple awardees to enhance statistical power and isolate the effects of different interventions by controlling for features that vary across interventions. This pooled analysis would likely focus on program effectiveness and the subdimensions of health, costs, and quality. Although it can add further insight into the performance of *individual* awardees, the main strength of a pooled analysis is to shed light on the effectiveness of certain *types* of interventions

and how that effectiveness is influenced by other factors, such as setting, context, or populations involved in the intervention. The strength of the analysis depends on the availability of suitable control populations and standardized and timely data on the individual interventions. The pooled analysis is designed to identify key elements of implementation effectiveness by taking advantage of larger sampler sizes and comprehensive analytic techniques.

Component 3: A Decision Strategy. The qualitative and quantitative comparisons and syntheses in Component 1 will address opportunities for cross-awardee learning in each of the five dimensions presented above. The pooled analyses from Component 2 will focus on program effectiveness and its subdimensions of health, costs, and quality, taking into account opportunities for pooling CMS, state, and other administrative data. A structured decision strategy would use data from these first two components to enable systematic consideration of key innovation features and outcomes to develop informed policy. The comparisons and syntheses that arise from pooled analyses have the potential for stronger internal and external validity of findings in the summary evaluation. These pooled analyses can thus be seen as an independent validation of findings from individual awardee, grouping, and Component 1 evaluations.

A summary evaluation may be carried out concurrently with the individual awardee and group evaluations. In order to accomplish this, the evaluators need to be coordinated in their work and have a clear plan for analysis, synthesis, and interpretation of their results.

Conclusion

The CMMI investment in new care and payment models is of potentially historic importance in terms of how we control health care costs while improving quality and outcomes. The evaluation of these awards will inform decisions about expanding the duration and scope of the models being tested. Despite the challenges, the evaluation and decision process must be of the highest technical quality, as well as transparent and well communicated. Thus, evaluators will have a critical role in the effort to reduce costs while maintaining quality in the delivery of health care. The strategy proposed in this report is put forward with these challenges in mind.

Acknowledgments

We wish to acknowledge the insightful guidance and input received from staff at CMS, including Timothy Day and Jan Valluzzi, both of whom provided thoughtful comments on several draft versions of the report.

Throughout this project, the HCIA Advisory Committee at RAND provided helpful guidance, comments, and advice on the development of the evaluation strategy. We wish in particular to acknowledge Rebecca Collins, Cheryl Damberg, Maria Orlando Edelen, John Engberg, Katherine Kahn, Ateev Mehrotra, Soeren Mattke, Deborah Saliba, Eric Schneider, and Bradley Stein.

The preparation of this document would not have been possible without the contributions of RAND staff, including communications analyst Kristin Leuschner, who assisted in developing and structuring the report and coordinating the contributions of our large team. Diane Bronowicz, Kerry Newsome, Steve Oshiro, and Nora Spiering also provided able assistance in the preparation and editing of this document.

Abbreviations

Abbreviation	Definition
ACE	angiotensin-converting enzyme
ADDQoL	Audit of Diabetes-Dependent Quality of Life
ADL	activity of daily living
ADR	adverse drug reaction
AHRQ	Agency for Healthcare Research and Quality
APC	activated protein C
ARB	angiotensin receptor blocker
BMI	body mass index
BP	blood pressure
BPRS	Brief Psychiatric Rating Scale
CAD	coronary artery disease
CAHPS	Consumer Assessment of Healthcare Providers and Systems
CG-CAHPS	Clinician and Group Consumer Assessment of Healthcare Providers and Systems
CHIP	Children's Health Insurance Program
CIPP	context-input-process-product
CMMI	Center for Medicare and Medicaid Innovation
CMS	Centers for Medicare & Medicaid Services
COOP	Dartmouth Primary Care Cooperative Information Project
COPD	chronic obstructive pulmonary disease
CV	cardiovascular
DD	difference-in-differences
EBM	evidence-based medicine
ED	emergency department
FACT	Functional Assessment of Cancer Therapy
FEV1	forced expiratory volume 1

Abbreviation	Definition
FS-ICU	Family Satisfaction in the Intensive Care Unit Survey
FTEs	full-time equivalent
GAF	Global Assessment of Functioning
HbA1c	glycosylated hemoglobin lab test
HCAHPS	Hospital Consumer Assessment of Healthcare Providers and Systems
HCIA	Health Care Innovation Award
HEDIS	Healthcare Effectiveness Data and Information Set
HIT	health information technology
HRQoL	health-related quality of life
IADL	instrumental activities of daily living
ICD-9	International Classification of Diseases, 9th Revision
ICU	intensive care unit
iEMR	interactive electronic medical record
IOM	Institute of Medicine
IT	information technology
IVD	in vitro diagnostic
LVSD	left ventricular systolic dysfunction
MCA	multi-criteria analysis
MEMO	Minimizing Errors/Maximizing Outcomes
MRSA	methicillin-resistant *Staphylococcus aureus*
MRSE	methicillin-resistant *Staphylococcus epidermidis*
NCQA	National Committee for Quality Assurance
NQF	National Quality Forum
NQMC	National Quality Measures Clearinghouse
PAM-13	Patient Activation Measure (short form)
PANSS	Positive And Negative Syndrome Scale
PHQ-9	Patient Health Questionnaire
PQI	Prevention Quality Indicator
PROMIS	Patient-Reported Outcomes Measurement Information System
QoL	quality of life
QOLI	Lehman's Quality of Life Interview

Abbreviation	Definition
ROI	return on investment
SIP	Sickness Impact Profile
SLDS	Satisfaction with Life Domains Scale
SOFI	Schizophrenia Objective Functioning Instrument
TA	technical assistance
VTE	venous thromboembolism

Chapter 1. Background, Evaluation Goals, and Overview

Background: Policy Context for the Health Care Innovation Awards

On November 14, 2011, the Center for Medicare and Medicaid Innovation (CMMI) within the Centers for Medicare & Medicaid Services (CMS) announced the Health Care Innovation Challenge. Through this initiative, CMS planned to award up to $900 million in Health Care Innovation Awards (HCIAs) for applicants who proposed compelling new models of service delivery or payment improvements that promise to deliver better health, better health care, and lower costs through improved quality of care for Medicare, Medicaid, and Children's Health Insurance Program (CHIP) enrollees. CMS was also interested in learning how new models would affect subpopulations of beneficiaries (e.g., dual eligibles, complex patients) who have unique characteristics or health care needs that could be related with poor outcomes (e.g., stinting on care). In addition, the initiative sought to identify new models of workforce development and deployment, as well as models that can be rapidly deployed and have the promise of sustainability. The Health Care Innovation Challenge was authorized in Section 1115A of the Social Security Act, which was added by Section 3021 of the Affordable Care Act (HHS/CMS, 2011).

The HCIAs provide an opportunity to implement and assess a broad range of innovative service delivery and payment models in local communities across the nation. The specific objectives of the Health Care Innovation Challenge were to

- "Engage a broad set of innovation partners to identify and test new care delivery and payment models that originate in the field and that produce better care, better health, and reduced cost through improvement for identified target populations.
- Identify new models of workforce development and deployment and related training and education that support new models either directly or through new infrastructure activities.
- Support innovators who can rapidly deploy care improvement models (within six months of award) through new ventures or expansion of existing efforts to new populations of patients, in conjunction (where possible) with other public and private sector partners." (HHS/CMS, 2011.)

In May and June 2012, CMS announced the recipients of the HCIAs. The 107 HCIA awardees proposed a broad range of new models of service delivery and payment improvements, with an emphasis on individuals with the greatest health care needs. Awards ranged from approximately $1 million to $30 million, and awardees include health care providers, payers, local government, public-private partnerships, and multipayer collaboratives (HHS/CMS, 2012).

CMS is engaging in a two-phase independent evaluation of HCIA awardees. In the first phase, CMS is conducting a scan of the awarded models and developing approaches (described in this report) for an independent evaluation of the awards. In the second phase, CMS will implement the independent evaluations following the approach outlined in the first phase. The evaluation is intended to identify successful models of care that CMS may wish to develop and disseminate further. CMS contracted with the RAND Corporation to conduct the first phase of the evaluation: the design of an independent evaluation strategy and the provision of evaluation technical assistance to awardees of the HCIAs.

An Evaluation Strategy for HCIA Awardees

The goal for the evaluation design process is to create standardized approaches for answering key research questions that can be customized to evaluate similar groups of awardees and that allow for rapid and comparable assessment across groups of awardees. The goal for the evaluation design is not to create a specific plan for each awardee. With such a large number of highly varied HCIA awardees, it would be cumbersome and prohibitively expensive to perform a separate, customized evaluation for each awardee. Further, the lack of common metrics inherent in such an approach would preclude comparisons of performance across awardees to identify common success factors, challenges, and unintended consequences of their efforts.

Goal of the Evaluation

The goal of the evaluation is to help CMS answer two key questions:

- Which kinds of innovative approaches result in reduced cost while improving or maintaining the standard of care, patient health and quality of life, and satisfaction of the workforce?
- To the extent that a particular approach is promising, what contextual factors need to be in place to make success likely, and what contextual factors might cause serious problems?

These two goals guide the selection of the evaluation strategy. The evaluations are designed first to focus on evaluating which interventions reduced costs while improving or maintaining health care and health care quality. The evaluations use additional dimensions related to program implementation, workforce issues, priority populations, and context to help CMS answer why a particular intervention was successful and how and in what settings it can be scaled up.

Challenges in Designing an Evaluation Strategy

RAND recognizes that there are complex challenges to designing an effective and comprehensive evaluation of the HCIA initiative. We summarize here a few of these challenges for evaluation design and implementation, which we address in the proposed strategy:

- **Evaluation skills of individual awardees.** Based on past experience with multiproject evaluations, we expect that the awardees funded under HCIA will vary widely in terms of their capacity to provide information for carrying out the external evaluation. In order to conduct the evaluation, the external evaluator will need to assess awardees' skill levels and establish a priority list if technical assistance becomes available.
- **Calculating program costs and cost savings.** A primary focus of program evaluators will be to identify models with the potential to reduce the total cost of care in ways that are financially sustainable for provider organizations. Key to achieving this aim is the ability of evaluators to measure effects of a comprehensive array of innovations on costs. The ability to calculate program costs and cost savings will be challenging for many awardees and has been addressed by technical assistance provided by CMS.
- **Awardee time burdens and coordination.** Both in this evaluation design contract and the subsequent contract to perform the evaluation, it is essential to balance the need for valid evaluation data against the desire to minimize time and resource burdens on individual awardees and patients. Our experience has demonstrated that it is preferable (whenever possible) for the evaluator to conduct secondary data analyses and to supplement these data with awardee interviews or additional materials directly from their projects. This approach enables the evaluator to obtain complete data and eases the evaluation burden on awardees.
- **Cross-cutting versus awardee-specific measures.** A crucial issue for the evaluation strategy is to achieve a balance between the use of measures that cut across awardees' efforts and those specific measures relevant to each innovation model being tested. The correct balance ensures that measures are general enough to inform cross-awardee comparisons while being relevant to a diverse array of awardee efforts and ensures that awardees do not bear an undue measurement burden.

Other sources of potential burden on the awardees are related to the structure of the HCIA program, which involves several contractors. Effective coordination among those contractors will enhance the efficiency with which each of them works with the awardees. We have attempted to address these and other challenges in designing the evaluation strategy.

Overview of the Document

This document represents the final draft of this report and incorporates revisions made in response to comments received from CMS on the previous drafts.

The remainder of the report consists of three chapters, as follows:

- In Chapter 2, we discuss strategies for evaluation at three levels: (1) evaluation of individual awardees, (2) evaluation within awardee groups, and (3) evaluation across all groups.
- In Chapter 3, we describe the general evaluation strategy, including the key evaluation dimensions, their importance for the evaluation, and the basic approach to measurement used for each of the evaluation dimensions. For each of the evaluation dimensions, we outline important subdimensions, evaluation questions, metrics, data, and approaches that may be used.
- In Chapter 4, we apply the evaluation dimensions, subdimensions, questions, metrics, data, and approaches to a summary evaluation strategy. The objective of this summary evaluation is to compare and synthesize findings from evaluations conducted at the awardee and group levels and from pooled analyses to assist in identifying those interventions that can be implemented more broadly, those that need testing in other settings, and those that may be deferred from current consideration for further investment.

Chapter 2. Strategies for Evaluation at Three Levels

The evaluation plan envisions that data collection and analysis will be carried out on three levels: at the level of the individual awardee, at the level of the awardee group, and as a summary evaluation that includes all awardees. In this chapter, we discuss strategies for evaluation at these three levels.

We expect that evaluations at all three levels will need to address several main questions, including the following:

- To what extent was the program implemented as designed? What modifications were made and why?
- What happened to cost of care—i.e., to what extent were awardee efforts associated with change in utilization or in the total cost of caring for target patients? Did utilization and costs decrease, increase, or stay the same? If there was change, how large was that change?
- What happened to quality of care, clinical indicators of outcomes, and patient health-related quality of life? Did quality of care and outcomes experienced by patients improve, decline, or stay the same? If there was change, how large was that change?
- How did workforce training and deployment change as a result of the program? What happened to levels of workforce satisfaction with the care they provide and their own jobs and working conditions?
- To what extent were there unintended positive or negative consequences from the program?
- What factors were critical to program implementation success? What was learned or developed as a result of the program? How can the program best be sustained and scaled up?

The ultimate goal is to learn from the approaches being tried in the HCIA interventions to identify strategies that can be employed widely to lower cost while improving care.

Evaluation of Individual Awardees

The first step at this evaluation level will be to obtain data directly from individual awardees. This may involve collection of

- documents, including program descriptions, plans, and projections, as well as materials, such as progress reports
- clinical and administrative data—e.g., data on patient characteristics, utilization, cost, and clinical outcomes

- self-report data from participating patients, including semi-structured qualitative interviews and structured surveys on patient experience and health outcomes
- self-report data from participating staff administering the program, including semi-structured qualitative interviews, structured surveys on staff experience in carrying out the program, and workforce outcomes
- observational data, such as observations of key program activities, ways in which staff are deployed in carrying them out, and other program features.

The specifics of how data will be collected from each awardee will be spelled out in the plans developed by external evaluators but will be based on the evaluation strategy. Evaluators may also make use of data already collected by CMS and contractors.

Evaluation Within Awardee Groups

CMS set out to identify groups of awardees with shared populations, interventions, outcomes, and settings that would allow for grouped evaluations. With the evaluation questions in mind, the RAND team advised CMS in this decision by considering several different grouping approaches:

- grouping by intervention type
- grouping by key outcomes
- grouping by "root cause"
- grouping by approach and treatment setting.

Each of these groupings has benefits, and we review them briefly below.

Grouping by intervention type. The first approach, grouping by intervention type, categorized the awardees by the type of intervention being implemented. While this approach was conceptually attractive, the eclectic intervention approaches across awardees may mean that within-group variation may be too large and between group variation too small for good comparisons to be made; also, most awardees are implementing multiple interventions.

Grouping by key outcomes. In the second approach, we developed nine outcome categories, including (1) improved health and quality of life, (2) improved access to care, (3) improved quality of care for specific conditions, (4) improved care for complex and high-risk patients, (5) reduced hospital admission and readmission, (6) decreased hospital length of stay, (7) reduced emergency department (ED) visits, (8) reduced unnecessary treatment or service, and (9) lower costs. Given the focus of the interventions on three key outcomes (i.e., better health care, better health, and lower costs through improvement), many of the awardees were in multiple categories. Specifically, all awardees were categorized in the "lower costs" and "improved care" categories. Thus, while this approach is useful from an evaluation perspective when thinking about potential measures, it is less useful for identifying unique groups of awardees.

Grouping by "root cause." The rationale for the third approach was that, by identifying the "root cause" that the intervention is addressing, we would be able to group awardees based on the problems they are addressing, regardless of the approach they use to accomplish that. We developed six primary categories of "root cause" problems, including (1) access to primary care (to include physician shortage, distance, wait time, language, culture, and disability), (2) access to specialty care (to include physician shortage, distance, and disability), (3) care transitions (to include acute care and home care or long-term care), (4) care integration (to include primary, specialty, acute, and behavioral health care), (5) care quality and variation (to include primary care, acute care, behavioral health, and specialty care), and (6) patient knowledge and decisionmaking. While these categories all describe important problems, the activities the awardees are undertaking typically address multiple root causes, and it would be difficult to associate outcomes with them. Further, drawing conclusions across different care settings would be challenging.

Grouping by approach and treatment setting. After considering these approaches, we proposed to group awardees on the basis of their approach *and* treatment setting. Among the awardees, there are several alternative approaches to the prevention of health crises among medically fragile patients in the community, thereby avoiding unnecessary ED use and hospitalizations, including readmissions. Some awardees focus interventions on medical diseases or patient complexity. Another group of awardees is pursuing hospital-based interventions, again taking different approaches. Others are focused on potentially high-value community-based interventions.

These categories suggest awardee groups with sufficient within-group homogeneity so that evaluations may answer important questions—for example:

- What are the most promising approaches to preventing health crises and limiting their impact?
- What are the most promising approaches to improving hospital care and transitions?

These approaches also lead to groups that have operational similarities in terms of care settings (hospitals, primary care, behavioral health care, long-term care), leading to similarities with the associated data collection and record systems. For each, there will be a need for specific measures to establish the extent to which they result in reduced or slowed growth in cost, improved quality of care, and improved health-related quality of life. Finally, they suggest ways of grouping awardees that are carrying out similar kinds of projects so that learning communities and technical assistance for measures, data collection, and other features can be provided in the most relevant ways.

Therefore, we suggest ten groups for awardee projects across three larger categories, as described below.

Management of Medically Fragile Patients in the Community

This category's approaches are characterized by a focus on high-risk patients who are receiving care primarily in community health care settings with the aim of reducing costs by improving care quality and reducing ED visits and hospital admissions. These approaches may guide their efforts in one of several ways:

1. Disease and condition-specific targeting (including but not limited to cardiac, asthma, dementia, diabetes, stroke, cancer, chronic pain, renal/dialysis)
2. Complex and high-risk patient targeting (multiple conditions, rural, low-income, end-stage)
3. Behavioral health patients being treated in community care settings.

Hospital Setting Interventions

This category's approaches are characterized by a focus on care of hospitalized patients with an aim of reducing the length of stay, intensity of utilization, and readmission. These include awardees who are targeting specific conditions, awardees who are targeting better use of different kinds of services (e.g., imaging), and awardees who are using such technology as data systems or remote monitoring to improve intensive care unit (ICU) care:

4. Condition-specific targeting (sepsis, delirium)
5. Acute care management
6. Improvement in ICU care, remote ICU monitoring.

Community Interventions

This category's approaches take place in community settings but are focused on aspects of care delivery rather than on categories of patients, though some may be focused on subgroups of patients. Some focus on better coordination and use of services, on health information technology, or on care management, while others are focused on primary care redesign, such as medical homes or the integration of depression care. Two other subgroups are focused on specific aspects of community care: medication management and shared decisionmaking. The groups in this category are

7. Community resource planning, prevention, and monitoring
8. Primary care redesign
9. Pharmacy and medication management
10. Shared decisionmaking.

Table 2.1 summarizes the ten groups.

Table 2.1. Summary of Awardees Categories and Groups

Category	Groups
Management of medically fragile patients in the community	Disease and condition-specific targeting (cardiac, asthma, dementia, diabetes, stroke, cancer, chronic pain, renal and dialysis) Complex and high-risk patient targeting (multiple conditions, rural, low income, advanced illness) Behavioral health patients being treated in community care settings
Hospital setting interventions	Condition-specific targeting (sepsis, delirium) Acute care management Improvement in ICU care, remote ICU monitoring
Community interventions	Community resource planning, prevention, and monitoring Primary care redesign Pharmacy and medication management Shared decisionmaking

Following discussions with CMS about the proposed groups and the assignment of awardees to the groups, RAND worked with CMS to finalize the assignment of awardees to the ten groups.

While these three types of approaches are designed to improve quality of care and reduce or slow the growth of cost through better care, they will do so in different ways and with different specific end points, and these differences will need to be taken into account in designing an evaluation plan. It will also be important to capture the specific structural features of programs (e.g., health information technology [HIT] improvements, workforce training, payment reform), the processes they include (e.g., care coordination, patient navigation, home visitation, care standardization), the effects on specific clinical outcomes and health-related quality of life, and the specific ways in which they are affecting cost in terms of reduced intensity of care, reduced ED visits, reduced hospitalizations and readmissions, and other factors.

In addition to this grouping structure, there are other characteristics that will be considered in the evaluation. These include

- target population characteristics (e.g., age, Medicare, Medicaid, CHIP)
- geographic characteristics, such as region and urban or rural setting
- program size, by funding level and number of beneficiaries
- workforce characteristics (e.g., type of staff, kinds of training and organization, and how types and levels are staff are deployed to undertake tasks within the system of care).

In Chapter 3, we describe an evaluation design that can be applied to each of these three groups, with modifications to account for between-group differences. These designs can serve as the basis for further development of awardee and group evaluation protocols.

Summary Evaluation

The value of a summary evaluation is the opportunity for CMS to examine interventions in the context of approaches and conditions that may not have been apparent before the individual and grouping evaluations were underway.

In Chapter 4, we present approaches for a summary evaluation of awardees and groups. These include a meta-analytic approach where possible, based on a systematic review of characteristics, and pooled data analyses to identify intervention strategies that are most effective in reducing costs while improving quality of care. Finally, we present structured approaches for establishing consensus interpretations of awardee and group evaluations, as well as for arriving at decisions about which approaches are worth scaling up, which are worth studying further and which should be deferred from current consideration for further investment.

Chapter 3. Evaluation Dimensions, Measures, and Designs

In this chapter, we describe the overall evaluation strategy, including the key evaluation dimensions, their importance for the evaluation, and the basic approach to measurement used for each of the evaluation dimensions. We begin by outlining a conceptual framework for the evaluation.

Conceptual Framework

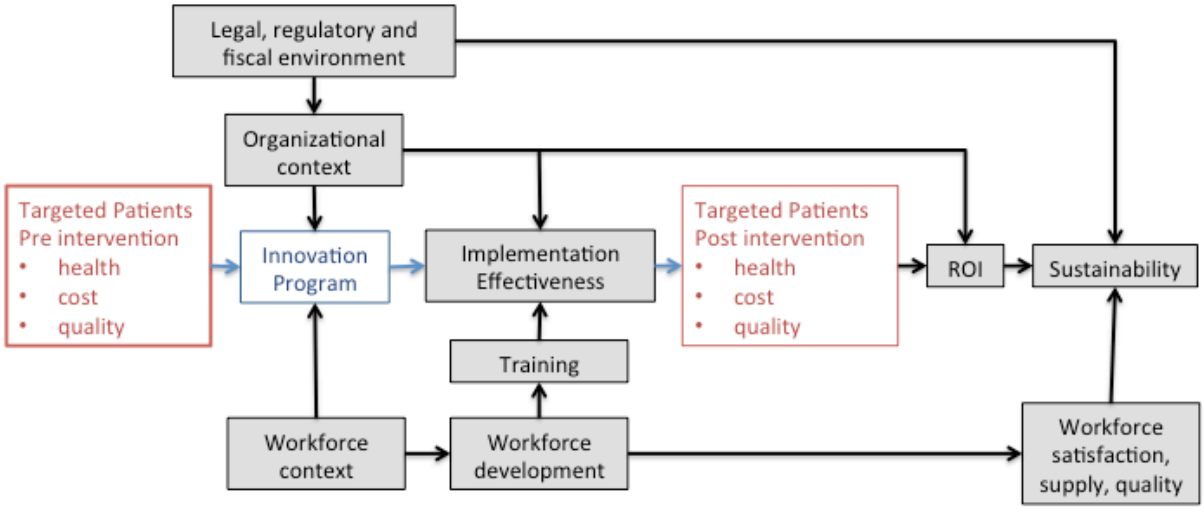
The conceptual framework for the evaluation is shown in Figure 3.1. The framework identifies key dimensions of the evaluation and indicates how they relate to a primary outcome of interest: the sustainability of an awardee program.

In the leftmost box, we depict the *health status and characteristics* of the target patient population. These characteristics motivate the *design of an innovation program*, which is also influenced by the *legal, regulatory, and fiscal environment*, the *organizational context*, and the *workforce context*.

The **implementation effectiveness** of a program is affected by *organizational context* and *workforce training* and can be measured along four dimensions: *program drivers* (i.e., the theory behind the program and intended drivers of change); *intervention components* (e.g., training, technical assistance), *dosage* (i.e., the "amount" of the intervention delivered to patients or the health system), and *fidelity* (i.e., adherence to planned procedures); and the reach of the program. **Program effectiveness** is characterized by the evaluation dimensions of *health*, *cost*, and *quality*. All of these factors affect the **return on investment (***ROI)*, which, along with *workforce satisfaction*, affects the overall **sustainability** of the program.

Each dimension in this framework represents a more complex set of elements, as we explain below. This framework is meant to be flexible so that it can be operationalized and interpreted by stakeholders with varying perspectives, including providers, evaluators, and CMS.

Figure 3.1. Conceptual Framework



Key Dimensions Within the Framework

In Table 3.1, we outline the key dimensions for the proposed evaluations. Below the table, we answer the following questions for each dimension:

- What is the dimension, and how is it defined?
- Why is it important for the HCIA project?
- How might we measure it?

This conceptual framework represents a modest evolution from the original framing of research questions outlined by CMS several months ago. In Appendix A, we offer a crosswalk of the framework and research questions. In Chapter 4, we apply these same evaluation dimensions to summary evaluation strategies and to a decision strategy for future investment, further study, or current deferral from future investments.

Table 3.1. Evaluation Dimensions

Category	Dimensions	Subdimensions
I. Implementation Effectiveness		
	A. Program drivers	1. Theory of change
		2. Theory of action
	B. Intervention	1. Components of the intervention
		2. Dosage
		3. Fidelity
		4. Self-monitoring
	C. Reach	1. Coverage
		2. Timeliness of implementation
		3. Secondary use of tools
II. Program Effectiveness		
	A. Health	1. Health outcomes
		2. HRQoL
	B. Costs	1. Program costs
		2. Utilization
		3. Expenditure
	C. Quality	1. Safety
		2. Clinical effectiveness
		3. Patient experience
		4. Timeliness
		5. Efficiency
		6. Care coordination
	D. Cross-cutting considerations	1. Equity and disparities
		2. Subgroup effects
		3. Spillover effects
III. Workforce Issues		
	A. Development and training	
	B. Deployment	
	C. Satisfaction	
IV. Impact on Priority Populations		
	A. Populations	1. Medical priority groups
		2. Nonmedical priority groups
	B. Impact	1. Cost reductions and savings
		2. Clinical outcomes
V. Context		
	A. Endogenous factors	1. Leadership
		2. Team characteristics
		3. Organizational characteristics
		4. Stakeholder engagement
	B Exogenous factors	1. Policy and political environment

NOTE: HRQoL = health-related quality of life.

Implementation Effectiveness

What Is Implementation Effectiveness?

Implementation effectiveness refers to the degree to which an intervention can be deployed successfully in real-world settings. We present three dimensions that are important in the evaluation of implementation effectiveness:

- **Program drivers:** the theory of change and theory of action underlying the intervention, or the drivers of change and mechanisms of action that will induce the intended change
- **Intervention components, dosage, fidelity, and self-monitoring:** the components of an intervention, its amount or "dosage," and the extent to which the innovation is carefully controlled or adapted to different contexts
- **Reach of the intervention:** the population reached by the intervention and the extent to which the implementation was timely, conducted as planned, and responsive to site-level constraints. Reach of the intervention also includes secondary use of tools—i.e., the extent to which any uses beyond the original design were discovered for HIT, decision support, and other intervention tools.

Why Is It Important to Measure Implementation Effectiveness?

The peer-reviewed research literature describes how implementation evaluations help to interpret ambiguous outcomes. For example, an innovation concept might be strong even though poor implementation has led to a null outcome. Without information on the extent and quality of implementation, decisionmakers might be tempted to reject essentially sound innovations. Alternatively, an intervention might be fundamentally weak even though positive outcomes have been achieved through the heroic efforts of strong implementers. Here, the absence of information on implementation might lead decisionmakers to adopt an ineffective innovation that cannot be translated or scaled up to other settings. In the late 1960s, these considerations prompted the evaluation field to begin developing models of implementation. For example, Stufflebeam's CIPP model (context-input-process-product) (Stufflebeam, 1983) was one of the first formal articulations of the need to look beyond outcomes, and this model is still used today by the Agency for Healthcare Research and Quality (AHRQ) in its evaluations of patient safety programs.

We propose that information on implementation effectiveness be used not to change the results of the program effectiveness evaluation, but rather to inform rapid evaluation, close-out, and scale-up of successful interventions. One of the goals of the HCIA program is to support innovators who can deploy programs quickly, within six months of the award. Thus, speed to implementation was a key consideration in the selection of awardees. CMS expected that models

would either already be operational and capable of rapid expansion or sufficiently developed so that they could be implemented rapidly. Moreover, a key goal of the HCIA program is to identify innovations that can be rapidly deployed more widely once they have been determined to be effective.

How Can We Measure Implementation Effectiveness?

We describe below how each of the three dimensions of implementation effectiveness can be measured.

Program drivers. Program drivers include the *theory of change* and the *theory of action* behind the intervention (Funnel, 2011). A theory of change identifies the mechanisms that catalyze or otherwise cause changes in individual and organizational behavior (note that this goes beyond "activities" such as training). A theory of action identifies the specific activities used to deliver the innovation. A theory of change, for example, might hold that physicians respond to certain new types of information, while a theory of action would describe how the new information might be produced or delivered. Drivers and related elements can typically be described through a review of program documents, literature on similar innovations, and interviews with program developers. HCIA awardees have developed diagrams that explain their program drivers. Table 3.2 describes research questions, metrics, data, and an approach that can be deployed to measure program drivers.

Table 3.2. Measurement of Program Drivers

Research Questions	Core Metrics	Data Sources	Analytic Approach
Theory of Change			
What are the central processes or drivers in the innovation by which change in behavior and systems is supposed to come about?	Program driver	Review of program documents, literature on similar innovations, and interviews with program developers	Descriptive
What implementation mechanisms are designed to activate the innovation's theory of change?	List of key implementation mechanisms associated with the innovation	Review of program documents, literature on similar innovations, and interviews with program developers	Descriptive
Theory of Action			
What are the central processes or drivers in the innovation by which patient or system-level action is meant to come about?	Program driver	Review of program documents, literature on similar innovations, and interviews with program developers	Descriptive
What implementation activities are designed to activate the innovation's theory of action?	List of key implementation activities associated with the innovation	Review of program documents, literature on similar innovations, and interviews with program developers	Descriptive

Intervention components, dosage, fidelity, and self-monitoring. Key measurement dimensions of the intervention include its components, quantitative indicators of dosage and fidelity, and self-monitoring activities of programs.

Components of the intervention refers to the specific activities by which the program seeks to induce better health outcomes at lower cost. For instance, some awardees have elected to use training programs, patient navigators, and shared decisionmaking as key components of their programs, while others have elected to implement medical homes, HIT, and new staffing. Overviews of awardee interventions for each group are shown in Table 4.1.

Dosage refers to how much of the innovation a health system or patient gets. This is perhaps the most straightforward aspect of implementation to measure. Examples include the intensity of care management interventions (e.g., phone, in person, frequency) and the amount of provider shared savings in a medical home. It is also desirable to include receipt and enactment of innovations, although measurement of these might not be feasible for every intervention type. *Receipt* is the extent to which the treatment, intervention, program, or innovation was actually noticed and comprehended by target populations, while *enactment* is the extent to which target populations have actually employed the innovation in use contexts (Schulte et al., 2009). *Fidelity* refers to how faithfully the innovation or program was delivered. This is often described as "adherence" to protocol (e.g., percentage of key elements included). Fidelity is typically measured through (a) self-report by providers, (b) surveys of the target population, or (c) judgment by trained observers, sometimes using formal rubrics. Schulte et al. (2009) suggests using program theory to flesh out key attributes of fidelity for a given intervention-context pair.

There is debate about the proper mix of fidelity and adaptation in good implementation. One camp of implementation researchers emphasizes the importance of fidelity and adherence to protocol, believing that the deviations dilute the causal impact of programs (Blakely et al., 1987; Bodilly et al., 2004). Others argue that skillful adaptation of programs is required to address different needs of subpopulations; encourage and maintain staff engagement and professionalism; develop and maintain community and target population buy-in; and address budgetary, resources, and other practical realities (see Mowbry et al., 2003; Barber et al., 2006). Blakely et al. argued that both views are right and proposed a "contingency theory" of implementation, according to which "highly specified innovations may require fidelity-supportive approaches, whereas broad policy innovations imply the use of decentralized strategies" (Blakely et al., 1987). One hypothesis is that the need for adaptation may increase with scale-up, as variation in context increases and as the intervention moves from highly motivated early adopters to a broad set of implementers and target populations.

Table 3.3 describes research questions, metrics, data, and an approach that can be deployed to measure intervention components, dosage, fidelity, and self-monitoring activities.

Table 3.3. Measurement of Intervention Components, Dosage, Fidelity, and Self-Monitoring

Research Questions	Core Metrics	Data Sources	Analytic Approach
Components			
What intervention components (e.g., training and technical assistance) were provided in support of implementation?	Description of components	Program proposals and reports	Descriptive
How much of each component was provided?	Quantitative indicators (e.g., technical assistance [TA] providers, training sessions)	Program reports	Descriptive
To what extent were the components available on an ongoing basis?	Quantitative indicators	Program reports, observations, surveys	Descriptive
How did unexpected events support or conflict with successful implementation of the innovation?	List of unexpected events and key impacts on implementation	Surveys—staff, surveys—clinicians, structured interviews with staff, structured interviews with clinicians	Qualitative analysis
Dosage			
What "dosage" of the innovation was *delivered* to patients, providers, and other target populations?	Quantitative measures	Program reports, observation, surveys	Descriptive
Fidelity			
In what ways is the innovation intended to be customized to specific use contexts?	List of ways in which the innovation is designed to be customized to the use context	Program proposals, administrative reports, interviews with staff, case studies	Descriptive
To what extent were systems in place to monitor implementation on an ongoing basis?	Implementation monitoring system	Program proposals, administrative reports, interviews with staff, case studies	Descriptive
How well did providers and sites adhere to planned procedures (including, as appropriate, procedures for customization)?	Adherence measures	Program proposals, administrative reports, interviews with staff, case studies	Comparative analysis of implementation activities relative to plans/ protocols
To what extent were the innovation and its components properly understood and used by target populations?	Self-reported ratings of understanding	Surveys	Descriptive

Research Questions	Core Metrics	Data Sources	Analytic Approach
Self-Monitoring			
What changes were made in response to self-monitoring?	Self-evaluation findings	Written reports and interviews with staff	Qualitative

Reach. Reach can be measured through three major subdimensions: the extent of the intervention's coverage (i.e., geographic reach, target population, number of individuals, organizations, or other units covered); the timeliness of its implementation; and the secondary use of tools that it generates. Program databases and administrative data can be used to measure reach, while activity logs and interviews can be used to evaluate the implementation timeline and timeliness. Reach can also be measured as the use of program tools in settings beyond those initially targeted; evaluations should therefore assess what secondary uses, if any, were discovered for HIT, decision support, and other intervention tools. Such data will help inform the ways in which secondary uses might be exploited to enhance benefits of the intervention(s) in other settings. Qualitative data from semi-structured interviews with key program staff, as well as CMS and awardee administrative data, might be used to assess the secondary use of tools. Table 3.4 describes research questions, metrics, data, and an approach that can be deployed to measure intervention reach through the subdimensions of coverage, timeliness, and secondary use of tools.

Table 3.4. Measurement of Intervention Reach

Research Questions	Core Metrics	Data Sources	Analytic Approach
Coverage			
What was the target population (e.g., patients, providers) after implementation?	Patient and provider characteristics	Awardee database, administrative records	Descriptive
How many patients, providers were reached?	Quantitative indicators • Number of providers engaged • Number of patients treated	Awardee database, administrative records	Descriptive
Timeliness			
To what extent was implementation timely, conducted as planned, and responsive to site-level constraints?	Implementation timeline	Activity logs and timelines, administrative records, interviews with staff	Descriptive
Secondary use of tools			
What secondary uses, if any, were discovered for IT, decision support, and other intervention tools?	Reports on secondary use of tools	Qualitative data from semi-structured interviews with key program staff	Descriptive—Use of intervention tools
How could secondary uses be exploited to enhance benefits of the intervention(s) in other settings?	Proposed methods to spread secondary uses to all participating providers	Qualitative data from semi-structured interviews with key program staff	Descriptive

Program Effectiveness

What Is Program Effectiveness?

Program effectiveness refers to assessments of an intervention's impact on outcomes of interest, referring to the aim of reducing cost through better care and better health. We present three outcome dimensions that are of interest in health care innovation:

Health. This dimension focuses on the impact of the intervention on health outcomes, including mortality, morbidity, health-related quality of life (HRQoL), and patient-reported outcomes.

Costs. This dimension focuses on program costs, impact on utilization, and expenditures resulting from the intervention.

Quality. This dimension focuses on improvements in care along several domains of quality: (1) safety, (2) clinical effectiveness, (3) patient experience, (4) timeliness, (5) efficiency, and (6) care coordination. The first five of these are based on five of the six domains of quality identified

by the landmark Institute of Medicine (IOM) report (NRC, 2001). The sixth IOM domain—equity—is discussed later in this evaluation framework as part of disparities, a cross-cutting consideration that applies not only to quality but also to access, health, and cost outcomes. The sixth subdimension in our framework, care coordination, includes performance measures, such as those tracked by HEDIS, and patient activation measures.

Cross-cutting considerations. We also discuss considerations that cut across the other dimensions in this section—health, costs, and quality. These considerations include equity and health care disparities issues, effects on specific subgroups of interest, and spillover effects.

Why Is It Important to Measure Program Effectiveness?

Evaluating program effectiveness is central to the aims of the HCIA program and critical to understanding whether an intervention delivers its intended effect. HCIA interventions are intended to "engage a broad set of innovation partners to identify and test new care delivery and payment models that originate in the field and that produce better care, better health, and reduced cost through improvement for identified target populations" (CMS, 2012). The HCIA awardees are expected to assess cost savings and to document improvements in health outcomes and quality over the three-year term of the award. They are also asked to project the intervention's effectiveness on an annualized basis after the term is finished.

Health. Improvements in health, health-related quality of life, and patient experience are a primary aim of most innovations in our health system. Documenting these improvements is a key part of establishing the value of interventions.

Costs. In order for the HCIA interventions to stimulate sustainable reductions in medical care costs, awardees must demonstrate cost savings for payers (i.e., CMS, state governments, and commercial health insurers) in a way that is financially viable for providers, while maintaining and promoting the health of patients. Overall, HCIAs generate savings when they reduce payer expenditures for patients who received the intervention, compared to payer expenditures for patients who did not receive the intervention, all else being equal. Interventions are financially viable for providers when they generate income sufficient to offset program implementation and operating costs. HCIA interventions that make ideal candidates for further development and dissemination are those that generate savings both for payers and providers while maintaining or improving patient health and quality of care. In addition, some HCIA interventions may have the potential to reduce the total costs of care at the national level—i.e., the costs for all payers combined. This level of cost reduction would be an example of "bending the cost curve" in health care.

Quality of care. Interventions should be safe, effective, patient-centered, timely, and efficient, and care should be coordinated across all sites and providers.

Cross-cutting considerations. Each of these subdimensions may have applications to evaluations of health, cost, and quality outcomes, and as such, they help evaluators explain *for whom* the intervention improves health, reduces costs, and maintains quality.

How Can We Measure Program Effectiveness?

To understand whether an individual HCIA generates savings, improves health, and maintains quality, we recommend that impact evaluations be conducted in three stages.

1. In the first stage, the evaluator will measure the summary impact of the individual intervention on pre-specified cost centers or events (such as hospital readmissions and emergency utilization), and on pre-specified components of the intervention.
2. In the second stage, the evaluator will assess the financial *viability of the intervention from the perspective of provider organizations*.
3. In the final stage, the evaluator will identify and describe *regulatory and payment policy changes needed, if any, to make the HCIA sustainable*.

We recommend that, wherever possible, evaluators use a common measurement and analytic framework to measure the health, cost, and quality impacts of HCIAs on payers at each of the three levels of analysis. The use of common measures and an analytic approach will support valid comparisons of individual programs, estimates of the magnitude of impact across awardees, and estimates of the relationships between design features and patient characteristics on program outcomes.

Health. This dimension can be measured along the subdimensions of health outcomes and HRQoL. These can be assessed using surveys, administrative data, and qualitative interviews. Where possible, evaluators might also consider causal pathways that explain observed relationships between specific components of the intervention and health outcomes. High-quality evaluations will have to address challenges associated with assuring comparability and quality of measures, managing measurement activities, minimizing survey and measurement burden on individual awardees, and establishing reasonable crosswalks from health measurement to health. A critical qualification of the evaluator will include demonstrated expertise in these addressing these challenges. Table 3.5 presents illustrative measures; the evaluators will have to determine the proper application of these or similar measures in the context of individual awards and awardee groupings.

Health outcomes. Measurement of health outcomes focuses on the extent to which the intervention improves desired health outcomes, including those health outcomes that are most important to the target population. Many health outcome measures are of interest, including mortality, morbidity, and functional health status.

Health-related quality of life. HRQoL is a multidimensional construct that incorporates various aspects of an individual's well-being. HRQoL is concerned specifically with the impact of a perceived health state on an individual's ability to live a fulfilling life (Bullinger et al., 1993). HRQoL incorporates physical, mental, and social well-being and includes positive as well as negative aspects of life and well-being.

There are several commonly used generic scales for measuring HRQoL, including SF-36, which provides scores on eight health domains and on two broad areas of subjective well-being: physical and mental health (Ware et al., 1995). The SF-36 is very widely used and has been extensively validated (Ware and Sherbourne, 1992; Shiely et al., 1997). Shorter versions of the SF-36 have been developed, and these can help to reduce survey burden in complex evaluation programs. These include the SF-12 (Burdine et al., 2000), the SF-8 (Turner-Bowker et al., 2003), and the SF-6D (Walter and Brazier, 2003).

Many disease-specific scales are also available, such as the Audit of Diabetes-Dependent Quality of Life (ADDQoL) for diabetes. The argument for using disease-specific quality of life (QoL) assessments is that they are better able to detect small changes in health status for people with the disease. Because these scales focus only on those domains that are thought to be affected by the disease, they often allow for a more comprehensive assessment in those specific domains. However, they do not permit comparison across disease conditions. Table 3.5 describes research questions, metrics, data, and an approach that can be deployed to measure health outcomes and HRQoL.

Table 3.5. Measurement of Health

Research Questions	Core Metrics (will vary by intervention)	Data Sources	Analytic Approach
Health Outcomes			
To what extent does the intervention improve desired health outcomes? Does the intervention result in any unanticipated negative health outcomes? Does the intervention affect health outcomes that are most important to the target population? Can we learn anything about causal pathways? In particular, for interventions with multiple components, which aspects of the intervention are primarily responsible for observed effects?	**Cross-Cutting** **Physical Health** • All-cause mortality • Complication rates • General health (Dartmouth Primary Care Cooperative Information Project [COOP]; Sickness Impact Profile [SIP]) • Functional health (example scales include the Katz ADL Index, Functional Independence Measure, Lawton IADL scale) **Mental and Behavioral Health** • Symptom rating scales (Brief Psychiatric Rating Scale [BPRS]) • Functional assessments (Global Assessment of Functioning [GAF]) **Disease and Condition-Specific** • Disease-specific mortality • Disease-specific complications • Asthma (symptom-free days, FEV1) • Diabetes (HbA1c, cholesterol, complications) • Depression—PHQ-9 • Schizophrenia (Positive And Negative Syndrome Scale [PANSS]; Schizophrenia Objective Functioning Instrument [SOFI]) • Substance abuse (Addiction Severity Index) • Stroke (functional assessments) • Cancer (Functional Assessment of Cancer Therapy [FACT]) • ICU complications (e.g., ventilator-associated pneumonia, thromboembolic events) • Pain (pain rating scales—e.g., visual analog rating scale, McGill Pain Questionnaire) • Physiologic measures (BP, BMI, liver function tests, drug screens) • Adverse drug events **Others** • Measures of physical health related to condition of interest, including comorbidities • Behavior change (medication adherence, substance use) • Patient characteristics—age, sex, race, zip code, comorbidities, etc. • Program characteristics • Provider characteristics • Site characteristics	Patient survey Clinical and administrative data Qualitative interviews with patients, caregivers, and health professionals	Interrupted time series Difference-in-difference analysis Matched comparison Thematic analysis of qualitative data Comparative case study (where applicable)

Research Questions	Core Metrics (will vary by intervention)	Data Sources	Analytic Approach
Health-Related Quality of Life (HRQoL)			
To what extent does the intervention improve quality of life? Can we learn anything about causal pathways? In particular, for interventions with multiple components, which aspects of the intervention are primarily responsible for observed effects?	Generic tools, such as the SF-36 or SF-12 Disease-specific tools, such as • ADDQoL for Diabetes • Lehman's Quality of Life Interview (QOLI) • Satisfaction with Life Domains Scale (SLDS) Patient characteristics—age, sex, race, ZIP code, comorbidities, etc.	Patient survey (e.g. ,SF-12, disease-specific instruments) Qualitative interviews with patients, caregivers, and health professionals	Interrupted time series Difference-in-difference analysis Matched comparison Thematic analysis of qualitative data

Cost. The purpose of the HCIA initiative is to test whether it is possible for CMS to "invest" money in new and innovative programs to generate care improvements and save costs. In order for programs to be sustainable over the long run, providers must offset program costs with increased revenue or reduced operational costs. Evaluators will assess the potential of awardees to generate sustainable cost savings by comparing program costs with total cost savings in calculating a program's return on CMS's investment. To this end, three subdimensions of cost should be evaluated: program costs, utilization, and expenditures. We group utilization and expenditures together in the cost section of our framework because they will be closely linked for any evaluator charged with estimating cost savings from HCIA.

Program costs. Program costs are calculated by creating an inventory of program inputs and calculating their costs (e.g., capital equipment, personnel, facilities, staff training, salaries, supplies, Internet access). This definition excludes the cost of care provided to patients that is reimbursable by health insurers. Program costs are covered by a combination of CMMI grants and in-kind contributions, with the relative share of each varying from program to program. The cost of many program inputs will be straightforward to quantify, such as the salaries of full-time staff. On the other hand, allocating the cost of one-time equipment purchases across time or attributing the cost of lost productivity from providers partly involved in the intervention can be difficult. Additionally, full accounting for program costs may require offsets of costs with savings if the program improves efficiency, even if it does not change utilization or improve health care and health. Programs that improve health and do not increase payer outlays but do cost the provider more must be evaluated carefully. Increases in provider costs are an important factor in determining whether and how to plan for sustaining and scaling up an innovation.

Utilization of health care services. Measures of utilization examine the extent to which quantities and rates of appropriate and inappropriate care have changed and whether there are any unintended consequences of these changes. Measures include utilization of specific services or categories of services, and intensity of care and indicators of appropriate and inappropriate care (e.g., services recommended by the U.S. Preventive Service Task Force).

Expenditures for health care services. The cost of a given service component from the payer's point of view (e.g., an inpatient stay or prescription medication) is the product of utilization and payment for one unit of the service, typically specified in fixed or negotiated fee schedules (of course, in capitation or even shared-savings environments, costs to the payer are different, but ultimately capitation or other rates reflect use of individual services). Summing the cost of individual service components yields a measure of total cost of care. Costs measured in this way are equivalent to medical care expenditures. In measuring cost of care, evaluators may choose to risk-adjust costs as appropriate to ensure that changes in cost over time reflect changes in resource use and not patient health status.

Table 3.6 describes research questions, metrics, data, and an approach that can be deployed to measure cost through the three subdimensions of program costs, utilization, and expenditures.

Table 3.6. Measurement of Cost

Research Questions	Core Metrics	Data Sources	Analytic Approach
Program Costs			
What were the fixed costs associated with program start-up? What are the variable costs associated with program operation? What are the anticipated new fixed costs associated with program sustainability?	Fixed costs (e.g., capital equipment, personnel, facilities, training needed to initiate services) Variable costs (e.g., salaries, supplies, Internet access)	Program proposals Awardee financial and progress reports	Descriptive
Utilization			
To what extent have levels of appropriate and inappropriate utilization changed? To what extent were there any unintended consequences for utilization? To what extent have levels of ED utilization changed? To what extent have rates of hospitalization and rehospitalization changed? To what extent has intensity of inpatient utilization changed?	ICD-9—total utilization of inpatient services and utilization of "service basket" categories CPT—total utilization of outpatient services and procedures Indicators of appropriate and inappropriate care (e.g., services recommended by the U.S. Preventive Service Task Force)	Administrative claims data Medicare and Medicaid charge data Awardee clinical record systems	Difference-in-differences (DD) (when sample size permits and a credible control group is available) Interrupted time series (when a credible control group is not available)
Expenditures for Health Care Services			
How are the models designed to reduce expenditures (e.g., changing the service the population utilizes, reducing the volume or utilization of services, changing the cost of services)?	Inventory of the "mechanisms of action" intended to reduce costs (e.g., adding a cost-beneficial service, reducing unnecessary care substituting less expensive input)	Program proposals Awardee financial and progress reports	Descriptive

Research Questions	Core Metrics	Data Sources	Analytic Approach
To what extent did the program change charges and expenditures for all care in the target population? To what extent did the program result in unintended charges and expenditures in the target population? To what extent do the models reduce or eliminate variations in charges or expenditures that are not attributable to differences in health status? What is the expected cost of sustaining these changes?	Utilization outcomes (from utilization evaluation dimension, above) Charges Cost to charge ratios (if needed) Program costs Program revenue Program savings Patient out-of-pocket costs Patient characteristics – age, sex, race, zip code, comorbidities, etc. Program characteristics Provider characteristics Site characteristics	Administrative claims data Medicare and Medicaid charge data Awardee financial reports—program cost, in-kind support, and revenue data	DD (when sample size permits and a credible control group is available) Interrupted time series (when a credible control group is not available) Return on investment (ROI) from the providers' perspective. (ROI = [(change in provider revenue due to the intervention + change in operating costs due to the intervention)/ program operating costs]

Quality. For the HCIA, quality is the extent to which the program provides better care that would plausibly lead to better health, which is conceptually consistent with CMS's first aim of better care, better health, and lower health care costs. In the proposed evaluation framework, programs can provide better care along six key subdimensions: safety, clinical effectiveness, patient experience, timeliness, efficiency, and care coordination. These subdimensions are closely aligned with the first five of IOM's six quality dimensions; the IOM domain of equity is addressed in the section of this chapter that is devoted to cross-cutting considerations. Although care coordination could be subsumed into other dimensions of quality—such as patient-experience or effectiveness—it features prominently in CMS's definition of "better care" and, as such, is highlighted separately in this section. Each awardee is expected to improve care across

some or all of these domains of quality and by selecting quality indicators and implementing a continuous improvement model of care delivery and evaluation. In addition to these efforts, clinical performance measures established by the Healthcare Effectiveness Data and Information Set (HEDIS) can be used to evaluate awardee impact on quality. These measures may address many domains, and most health care providers are experienced at tracking and reporting these data for accreditation purposes. Therefore, evaluators may be able to rely on awardee clinical databases to track performance quality. Additional sources of data include Medicare, Medicaid, and private payer administrative claims. To assess the impact of programs on these measures, an interrupted time series approach is recommended in most cases. Following is a discussion of the six subdimensions of quality that we recommend in HCIA evaluations.

Safety. Safety pertains to the avoidance of injuries and harm that may occur during the delivery of health care. *To Err is Human*, IOM's seminal work on patient safety (Institute of Medicine, 2000), suggested that as many as 98,000 deaths per year could be attributed to safety lapses. Key patient safety measures include medication reconciliation, fall risk assessments, and decubitus ulcer assessments. Patient safety can be measured using a number of data sources. For example, administrative claims can be used to estimate some indicators of safety in the hospital setting. Medical chart reviews are a key source of data for measuring patient safety. Table 3.7.1 describes research questions, metrics, data, and an approach that can be deployed to measure quality through the subdimension of safety.

Table 3.7.1. Measurement of Safety

Research Questions	Core Metrics (will vary by intervention)	Data Sources	Analytic Approach
Safety			
To what extent do the models improve patient safety?	**Inpatient** • Rates of surgical complications • Unanticipated reoperative rates • Pressure ulcer risk reassessment • Pressure ulcer rate • Fall risk management • Ventilator-associated pneumonia • Fall rate • Central line infections • Selected conditions without present-on-admission (POA) flag **Cross-cutting** • Error reporting • Medication reconciliation • Adverse drug reactions (ADRs) (AHRQ National Quality Measures Clearinghouse) • Screening for Future Fall Risk (NQF#0101)	Provider survey—Minimizing Errors/ Maximizing Outcomes (MEMO) Certified electronic health records Administra-tive claims data Incidence reports	Descriptive Analytic—interrupted time series DD (when credible control is available)

Clinical effectiveness. Clinical effectiveness refers to the extent to which care is consistent with the best scientific knowledge. The concept of effectiveness includes avoiding underuse of effective treatments and overuse of ineffective treatments. In this context, effectiveness may be thought of as the delivery of *key processes*, which are interventions performed on and for patients that can plausibly be associated with improved health. Unique measures of these processes are available for different care settings. For example, the assessment of hemoglobin A1c for patients with diabetes or the mammography rate for at-risk women are key process measures for physician practices, while the proportion of heart attack patients who were prescribed aspirin at discharge or the proportion of heart failure patients given a left ventricular function assessment are key process measures for acute care hospitalizations. Effectiveness can be measured using a number of data sources, including administrative claims and electronic health record data. Table 3.7.2 describes research questions, metrics, data, and an approach that can be deployed to measure quality through the subdimension of clinical effectiveness. Here again, these measures are illustrative; evaluators will have to determine the proper application of these or similar measures in the context of individual awards and awardee groupings.

Table 3.7.2. Measurement of Clinical Effectiveness

Research Questions	Core Metrics (will vary by intervention)	Data Sources	Analytic Approach
Clinical Effectiveness			
To what extent do the models improve the effectiveness of patient care? To what extent have clinical condition indicators changed? To what extent does the intervention affect key performance goals, such as compliance with treatment guidelines?	**<u>Outpatient</u>** Diabetes: • Eye exam, NQF # 0055 • Foot exam, NQF # 0056 • Urine protein screening, NQF # 0062 • Diabetic lipid and hemoglobin A1c profile • Proportion of persons with diabetes with an HbA1c value greater than 9 percent • Proportion of the diabetic population with an HbA1c value less than 7 percent • Lipid control among persons with diagnosed diabetes • Proportion of persons with diagnosed diabetes whose blood pressure is under control Asthma and chronic obstructive pulmonary disease (COPD): • Use of appropriate medications for asthma (NQF#0036) • Medication management for people with asthma (NQF#1799) • Asthma: pharmacologic therapy for persistent asthma (NQF#0047) • Management plan for people with asthma (NQF#0025) • COPD: spirometry evaluation (NQF#0091) • COPD: inhaled bronchodilator therapy (0102) Heart failure: • Beta-blocker therapy for left ventricular systolic dysfunction (NQF#0083) • Left ventricular ejection fraction assessment (NQF#0079)	Certified electronic health records Claims data (ICD-9s and CPT Category II codes) Registries Electronic reporting mechanisms Patient surveys (e.g., Patient Health Question-naire, PROMIS, Confusion Assessment Method)	Descriptive Analytic—interrupted time series DD (when credible control is available)

Research Questions	Core Metrics (will vary by intervention)	Data Sources	Analytic Approach
	Cardiovascular • Controlling high blood pressure (NQF # 0018) • Oral antiplatelet therapy prescribed for patients with coronary artery disease (CAD) (NQF #0067) • ACE inhibitor or ARB therapy–Diabetes and LVSD (NQF#0066) • Beta-blocker therapy (NQF #0070) • CAD: lipid control (NQF # 0074) • IVD: lipid control (NQF 0075) Cancer • Plan of care for pain (NQF#0383) • Pain intensity quantified (NQF#0384) • Proportion receiving chemotherapy in the last 14 days of life (NQF#0210) Pain • Self-reporting of pain (CMS, NQF #0676) • Self-reported measure of severe pain (CMS, NQF #677) Depression • Screening for clinical depression (NQF#0418) • Antidepressant medication Management (#0105) • Depression remission (NQF #0710) Other behavioral health: • Adherence to antipsychotic medications for individuals with schizophrenia (NQF#1937) • Adherence to antipsychotic medications for individuals with schizophrenia (NQF#1879) • Diabetes and cardiovascular (CV) disease screening and monitoring for people with schizophrenia or bipolar disorder (NQF#1932, 1927, 1933, 1934)		

Research Questions	Core Metrics (will vary by intervention)	Data Sources	Analytic Approach
	Prevention: • (A) Tobacco use assessment, (B) tobacco cessation intervention (NQF#0028) • Colorectal cancer screening (NQF#0034) • Cervical cancer screening (NQF#0032) • Adult weight screening and follow-up (NQF#0421) Immunizations: • Childhood immunization status (NQF#0038) • Influenza vaccination (NQF#0043) Wellness: • Well-child visits in the third, fourth, fifth, and sixth years of life (NQF#1516) • Body mass index (BMI) 2 through 18 years of age (NQF#0024) • Adult weight screening and follow-up (NQF#0421) • Lead paint screening in children (NQMC#007068) Neonatal: • Frequency of ongoing prenatal care (NQF#1391) Sexually transmitted infections: • Chlamydia screening (NQMC#007072) Medication adherence: • Proportion of days covered (NQF#0541)		

Research Questions	Core Metrics (will vary by intervention)	Data Sources	Analytic Approach
	Inpatient Mobility level assessment Pain assessment Assessment of activities of daily living (ADLs) Sepsis • Percentage of patients with severe sepsis/septic shock who had 2 sets of blood cultures collected within 24 hours following severe sepsis/septic shock identification • Percentage of eligible patients with severe sepsis or septic shock who were started on activated protein C (APC) within 48 hours following severe sepsis/septic shock identification • Percentage of patients with severe sepsis/septic shock who were assessed for activated protein C (APC) eligibility within 24 hours following severe sepsis/septic shock identification • Percentage of patients with severe sepsis/septic shock and an organism other than MRSA or MRSE who had Vancomycin (or Linezolid) discontinued within 72 hours following severe sepsis/septic shock identification • Percentage of patients with severe sepsis/septic shock who received a recommended broad-spectrum antibiotic within 24 hours following severe sepsis/septic shock identification Percentage of patients with severe sepsis/septic shock who received Vancomycin (or Linezolid) within 24 hours following severe sepsis/septic shock identification		

Research Questions	Core Metrics (will vary by intervention)	Data Sources	Analytic Approach
	Delirium • Proportion of patients meeting diagnostic criteria on the Confusion Assessment Method (CAM) ICU • Proportion of patients meeting diagnostic criteria who are receiving ICU-level care • Percentage of patients who were transferred or admitted to the ICU within 24 hours of hospital arrival and who had blood cultures performed within 24 hours prior to or on the day prior to arrival, the day of arrival, or within 24 hours after arrival to the hospital • Percentage of patients who received venous thromboembolism (VTE) prophylaxis or have documentation why no VTE prophylaxis was given the day of or the day after initial admission (or transfer) to the ICU or surgery end date for surgeries that start the day of or the day after ICU admission (or transfer) **Cross-cutting** Self-reported outcomes		

Patient experience. Patient experience pertains to patients' perceptions of specific aspects of care they receive from providers. Patient experience is thought to be a useful patient-reported indicator of quality because it provides a clear basis for actionable improvements and is less dependent on patient preferences and expectations than patient satisfaction measures (Cleary et al., 1998). Key measures of patient experience include summary ratings for providers, as well as individual ratings for provider communication, access to care, care coordination, courtesy and helpfulness of providers, cultural competency, self-management education, and telehealth equipment and processes. A number of approaches can be used to assess patient experience, the strongest of which include patient surveys, focus groups, and interviews. Other approaches, such as the use of mystery shoppers and such online health care rating forums as WebMD and HealthGrades, are less scientifically valid. Of particular importance are the CAHPS surveys, a suite of standardized, validated tools to measure patients' experience with care that have been endorsed and adopted by CMS and used widely by hospitals and other providers. The various CAHPS surveys (e.g., HCAHPS for hospitals) ask patients to rate their care across a wide range of settings, including health plans, physician practices, behavioral health, hospitals, dialysis centers, nursing homes, and home health agencies. Table 3.7.3 describes research questions,

metrics, data, and an approach that can be deployed to measure quality through the subdimension of patient experience.

Table 3.7.3. Measurement of Patient Experience

Research Questions	Core Metrics	Data Sources	Analytic Approach
Patient Experience			
In what ways are aspects of patient experience (e.g., access, perceived care coordination, provider-patient communication) enhanced by the intervention(s)? In what ways are aspects of patient experience worsened by the intervention?	**Outpatient** Measures of patient experience with outpatient care (e.g., "health care provider showing respect [percentage], health care provider spending enough time with patient [percentage]") **Inpatient** Measures of patient experience with inpatient care (e.g., rating of provider communication, overall hospital rating, rating of hospital environment) Family members' satisfaction with care in the ICU **Cross-cutting** Measures of patient involvement in shared decisionmaking process	Patient survey (CG-CAHPS, HCAHPS, FS-ICU)	Descriptive Analytic—pre-post DD (when credible control is available)
To what extent does the intervention affect measures of patient activation?	Patient activation	Patient survey (PAM-13)	Descriptive Analytic—pre-post DD (when credible control is available)

Timeliness. Timeliness is considered by the IOM (IOM, 2001) to be the minimization of waiting times and avoidance of potentially harmful delays in care. Timeliness is an important measure of quality because getting care without delays improves patients' experience with care and increases the probability of maximizing patient outcomes. Timeliness has been applied in both inpatient and outpatient settings and can be measured in a number of ways. For example, timeliness may be measured as the proportion of patients in a practice who are able to schedule a nonurgent (routine) visit within four days. In the inpatient setting, timeliness might be measured as the time it takes to initiate recommended broad-spectrum antibiotics for patients with septic shock. Table 3.7.4 describes research questions, metrics, data, and an approach that can be deployed to measure quality through the subdimension of timeliness.

Table 3.7.4. Measurement of Timeliness

Research Questions	Core Metrics	Data Sources	Analytic Approach
Timeliness			
To what extent do the models improve the timeliness of care?	**<u>Outpatient</u>** Length of wait for an appointment for urgent care Length of wait for an appointment for nonurgent care Delaying needed care Median time from ED arrival to ED departure for discharged ED patients **<u>Inpatient</u>** Admit decision time to ED departure time for admitted patients Documentation of pressure ulcers within 24 hours Sepsis: • Median time to initiation of a recommended broad-spectrum antibiotic following severe sepsis/septic shock identification • ICU access and exit blocks	Patient surveys Provider surveys	Descriptive Analytic—pre-post DD (when credible control is available)

Efficiency. Efficiency refers to the extent to which care avoids waste and provides value to patients, payers, or purchasers. As health care costs escalate, the demand for efficiency will presumably grow. The measurement of efficiency is an emerging field in health care, and little consensus exists about the best and most appropriate measures. Efficiency has frequently been assessed using proxy utilization measures, such as the rate of readmissions or ambulatory care–sensitive admissions. Many researchers are working to develop effective cost measures and to combine them with quality measures to better assess the efficiency of care delivered to patients. However, this work is still in progress, and consensus on efficiency measures has yet to emerge. Table 3.7.5 describes research questions, metrics, data, and an approach that can be deployed to measure quality through the subdimension of efficiency.

Table 3.7.5. Measurement of Efficiency

Research Questions	Core Metrics	Data Sources	Analytic Approach
Efficiency			
To what extent do the models improve the efficiency of care?	**<u>Outpatient</u>** Emergency department visits (NCQA) Hospital admissions for ambulatory-sensitive conditions Utilization of MRI and CT Utilization of observation care and substitution of lower-intensity interventions Inpatient admissions **<u>Inpatient</u>** Length of stay Readmissions Utilization of observation care and substitution of lower-intensity interventions ED revisit	Claims data Certified electronic health record (HER)	Descriptive Analytic—pre-post DD (when credible control is available)

Care coordination. Care coordination has been described as the "deliberate organization of patient care activities between two or more participants involved in a patient's care (including the patient) to facilitate the appropriate delivery of health care services" (McDonald et al., 2007). Care coordination is widely viewed as a necessary response to the fragmentation of care that underlies many of the problems in the U.S. health system, including patient safety breakdowns, chronic care inadequacies, widespread inefficiencies, and escalating costs. Almost all quality improvement efforts today, including most of the HCIA innovations, aim to improve the coordination of care. Measures of care coordination may identify the extent to which providers link patients to community resources, provide a health care home, work as a team, support patients' self-management goals, and assess patients' needs and goals. Measuring care coordination may rely on primary data collected by provider or patient and caregiver feedback. Table 3.7.6 describes research questions, metrics, data, and an approach that can be deployed to measure quality through the subdimension of care coordination.

Table 3.7.6. Measurement of Care Coordination

Research Questions	Core Metrics	Data Sources	Analytic Approach
Care Coordination			
To what extent did the models improve care coordination?	**Cross-cutting** Timely transmission of transition record Transition record with specified elements received by discharge patients Timely communication of other information (e.g., upon referral) Post-discharge continuing care plan created Post-discharge continuing care plan transmitted to next level of care provider upon discharge Development and sharing of care plans among other settings Percentage of patients with care plan in record Advance care plan Patient reports on care coordination and transitions	Qualitative data from semi-structured interviews with key program staff Claims data Clinical records Provider survey Patient survey	Descriptive—inventory of care coordination processes Analytic—interrupted time series DD (when credible control is available)

Cross-cutting considerations. Several dimensions cut across health, cost, and quality outcomes, including disparities, subgroup effects, and spillover effects. Of interest is the relationship of these dimensions with variations in patient outcomes. Also of interest are the program characteristics that may be in the causal pathway between these dimensions and outcomes. Key program characteristics to assess include workflow redesign, HIT, telemedicine, care coordination, patient navigators, shared decisionmaking, and other aspects of the intervention.

Disparities. Another outcome to be measured is the effect of the intervention in reducing disparities on the basis of race, ethnicity, gender, age, etc., in patient enrollment, access, quality of care, and outcomes. Patient surveys, clinical data, administrative data, and qualitative interviews with patients and caregivers can be used to assess this cross-cutting dimension.

Subgroup effects. Another, related dimension is subgroup effects—i.e., for outcomes of interest (efficiency, value) for which a main effect *was not* detected, it is important to assess whether there was a subgroup of patients for whom an effect *was* detected. Conversely, for outcomes of interest for which a main effect *was* detected, it is important to assess whether there was a subgroup of patients for whom the effect *was weaker, stronger, or not detected*. These evaluations may focus on differences that are related to patient sociodemographic characteristics,

many of which are also covered in typical disparities evaluations. Subgroup evaluations, however, go a step further by assessing the whole distribution of care across patients and identifying any factors that are associated with the distribution. Evaluators should identify clinical, risk, socioeconomic, place, and other characteristics of patients, providers, and settings in which a subgroup effect is (or is not) detected. An understanding of subgroup effects can be used to identify those characteristics of patients and settings that have the greatest influence on the probability of a good outcome. This information can be helpful in targeting the intervention(s) in other settings.

Spillover effects. Interventions have the potential to produce outcomes (both positive and negative) that extend beyond the target population or specific area of focus. Therefore, it is important to identify any spillover effects from the intervention, whether at different sites, among providers, among non-targeted patients (through unintended effects on all services), or among targeted patients (through unintended utilization of other services). Evaluators may identify the program characteristics or factors that influenced these effects—e.g., the extent to which workflow redesign, HIT, telemedicine, and other structural aspects of the intervention result in spillover effects at the site(s) or among providers. The effect of contextual elements should also be considered—e.g., the extent to which care coordination, patient navigators, shared decisionmaking, and other aspects of the intervention(s) result in spillover effects among non-targeted patients.

Table 3.8 describes research questions, metrics, data, and an approach that can be deployed to measure cross-cutting considerations through the subdimensions of equity and disparities, subgroup effects, and spillover effects.

Table 3.8. Measurement of Cross-Cutting Considerations

Research Questions	Core Metrics	Data Sources	Analytic Approach
Equity and Disparities			
What contribution did the program make in reducing disparities in patient access to care? What contribution did the program make in reducing disparities in enrollment of targeted patients in intervention?	Access and enrollment outcome measures Characteristics of patients, programs, providers, and sites	Patient surveys Clinical and ad-ministrative data Qualitative interviews with patients, caregivers, and health professionals	Analytic—supplement access and enrollment outcomes analyses with subgroup analyses

Research Questions	Core Metrics	Data Sources	Analytic Approach
To what degree do the model(s) result in reductions in or elimination of disparities in quality of care?	Outcomes from access to care evaluation dimension, above Self-reported measures of access (e.g., inability to obtain or delay in obtaining medical care, inability to obtain or delay in obtaining prescription meds, usual primary care provider)	Provider surveys Patient surveys (e.g., CAHPS) http://healthindicators.gov/Indicators Administra-tive claims Awardee clinical data Registries	Descriptive Analytic—interrupted time series
To what degree does the program result in reductions in or elimination of disparities in patient outcomes?	• Mortality • Morbidity • Functional health status • Other health indicators • Patient-reported outcomes • Patient characteristics—age, sex, race, ZIP code, comorbidities, etc. • Program characteristics • Provider characteristics • Site characteristics	Patient survey Clinical and ad-ministrative data Qualitative interviews with patients, caregivers, and health professionals	Analytic—supplement main outcomes analyses with subgroup analysis
What program characteristics influenced reductions of disparities in access, quality, or outcomes?	Program characteristics • workflow redesign • HIT • telemedicine • care coordination • patient navigators • shared decisionmaking • other aspects of the intervention	Qualitative data from semi-structured interviews with key program staff Qualitative data from focus groups with target patients	Descriptive—factors influencing positive effects

Research Questions	Core Metrics	Data Sources	Analytic Approach
Subgroup Effects			
In outcomes of interest (health, costs, quality) for which a main effect was not detected, was there a subgroup in which an effect was detected? In outcomes of interest (health, costs, quality) for which a main effect was detected, was there a subgroup of patients for which the effect was stronger, weaker, or not detected?	Utilization, cost, and clinical outcome measures	Outcome distributions	Subgroup analysis
What were the characteristics of patients, providers, and settings in which a subgroup effect was detected?	Characteristics of patients, programs, providers, and sites	Awardee database, administra-tive records	Subgroup analysis
What characteristics of patients and settings influencing subgroup effects could be used to target the intervention(s) in other settings?	Characteristics of patients, programs, providers, and sites	Administra-tive records, interviews with key program staff	Subgroup analysis
Spillover Effects			
What, if any, were the positive and negative spillover effects of the intervention(s)? • At site(s) • Among providers • Among non-targeted patients (through unintended effects on all services) • Among targeted patients (through unintended utilization of other beneficial services)	Reported value and efficiency effects at sites, among providers, and among non-targeted patients Utilization of other services by targeted patients	Qualitative data from semi-structured interviews with key program staff CMS and awardee administrativ e data	Descriptive — inventory of unintended positive effects DD approach with matched controls

Research Questions	Core Metrics	Data Sources	Analytic Approach
What program characteristics and factors influenced these effects? • To what extent did workflow redesign, HIT, telemedicine, and other aspects of the intervention result in spillover effects at the site(s)? • To what extent did care coordination, patient navigators, shared decisionmaking, and other aspects of the intervention(s) result in spillover effects among non-targeted patients?	Program characteristics • Workflow redesign • HIT • Telemedicine • Care coordination • Patient navigators • Shared decisionmaking • Other aspects of the intervention	Qualitative data from semi-structured interviews with key program staff	Descriptive—factors influencing positive effects
How can spillover effects be exploited in future implementation efforts using similar models of care?	Proposed bundling of structural and care innovations to address multiple conditions and support multiple treatments	Qualitative data from semi-structured interviews with key program staff	Descriptive

Workforce Issues

What Are Workforce Issues?

There are three key types of workforce issues to be considered: development and training, deployment, and satisfaction.

Development and training. Staff education and training are key to successful implementations of innovations and system transformation. As a practice brings in the personnel and skills required for a successful implementation of its model, it is important to understand what works best: a training process and other strategies to add new skills to current workers or contracts with outside providers who already have those skills. Workforce issues of note for the HCIA program thus include identifying new roles for and retraining existing health professionals, identifying the skills needed to support health care innovations, and training new types of workers to provide nonclinical care. Workforce changes might also include exploring team-based approaches to better utilize an effective mix of health care practitioners.

Deployment. How workers are deployed and interact with patients is also critical to the success or effectiveness of many of the awardees' interventions. The frequency and mode by

which programs contact patients and other provider staff are related to the success of innovation programs (Peikes et al., 2009). Roles and tasks of members of the care team are also important—for effectiveness, provider efficiency, and satisfaction. A key issue in health care is what tasks should be performed by workers with what level of training. For example, physicians are encouraged to practice "at the top of their license." With this model in mind, physicians and other staff may become part of care teams with restructured responsibilities.

Satisfaction. Health care providers in the United States undergo distinct training pathways, occupy different but overlapping roles in health care delivery, and are separately licensed and credentialed. The largest professional categories include physicians (roughly 900,000 in the United States), registered nurses (more than 2.5 million, including nurse practitioners), pharmacists (275,000), and licensed practical and vocational nurses (750,000). Key to the types of interventions under way among the HCIAs are other personnel who often work on the front lines—as the first line of access with patients or performing key monitoring, coordination, and outreach tasks. These include medical assistants (500,000); nursing aides and orderlies (1.5 million); home health and personal care aides (1.8 million); and various types of social and community workers, aides, and other allied health personnel. Job satisfaction is key to providers' willingness to be part of this workforce and to their ability to perform their work effectively. It is also key to the smooth functioning of a provider organization.

Why Are Workforce Issues Important for the HCIA Project?

Transforming the health system requires transformation of the health workforce. A critical challenge of delivery system reform is to identify and test new ways to create the workforce of the future, a workforce that will deliver and support new care models. For example, a health innovation might require nurses to provide care coordination in primary care settings, or ask new community-health workers to serve as a bridge between the health care system and the patients, or ask community-based teams of practitioners to provide clinical care and intensive care management services for the most complex patients. In selecting HCIA awardees, CMS sought to identify programs that demonstrate the ability to rapidly develop and deploy individuals who are capable of taking on new and expanded roles and that encourage collaboration among educational institutions, health care practitioners, and delivery systems. Health care re-engineering initiatives also require skilled individuals who are capable of documenting current processes and helping to design and implement new processes. Also important is the ability of the organization to build workforce capacity and capability and to demonstrate the potential for scale-up of effective interventions. Finally, change may require subtle but important cultural shifts in how organizations are managed, how workers interact, and how they perceive their roles in relation to each other and to patients.

How Might Workforce Issues Be Measured?

There are several options for measuring these workforce dimensions:

Development and training. Key elements to be measured include the extent to which programs provide training to use existing staff and incorporate new kinds of staff effectively; the level of investment in training required to fill workforce gaps, and; the effectiveness and efficiency of various training models. Awardees are already required as part of their quarterly reporting to report counts and full-time equivalents (FTEs) of staff involved in the innovation. Providers may also require specific data or feedback from patient experience, utilization, or other data so that they can respond and improve their processes and performance. Comparative information about their performance relative to that of other providers could prove particularly helpful. Assessing the extent to which the workforce is retained will also be important. Although broadly generalizable benchmarking data will not be available to compare how well different awardees are able to recruit and retain staff, it may be reasonable to compare rates of retention and turnover for the innovation staff to normal rates of retention and turnover for others employed by the organization. Large differences in retention are likely signs of low satisfaction with some aspect of the innovation, though evaluators may not be able to identify specific aspects of the intervention that caused the dissatisfaction—at least not without interview or survey data from staff.

Deployment. Considering that many of the innovations create new types of teams and new types of collaborations, it may be important to evaluate awardees based on measures of "teamness"—i.e., how well these newly created groups function. An additional issue that may be useful in considering workforce deployment is also how workforces are *utilized* in the innovation and how their roles may or may not differ from their traditional role(s). This may involve not only the number of individuals needed, but also the types of services provided by the innovation staff, whether or not staff members are able to practice "at the top of their license," how the innovation proposes to change the workforce or the care processes, how the legal and regulatory climate affect which staff can perform what tasks in relations to patients, and how policy can influence the diffusion of staff deployment innovations to other providers and locations. As noted above, it is important to capture how providers contact patients across the full range of possibilities (in groups, in person, by phone, one-on-one) and what is most effective and efficient. It is also important to measure how the innovation has changed the incidence of stress or burnout among staff and to determine whether rates of staff retention and turnover have changed over the course of the intervention. This will affect the supply of providers and staff available for the intervention and has an impact on the ultimate sustainability of the intervention.

Satisfaction. To understand staff satisfaction, it is important to measure the extent to which different kinds and levels of staff are satisfied or dissatisfied with the care they are able to provide. Assessment of staff satisfaction with working conditions is also important. This would include such factors as satisfaction with colleagues, other staff, income, organizational policies, etc. Other areas to measure include the extent to which different kinds and levels of staff report satisfaction or dissatisfaction with specific components of the intervention. This would include components introduced as part of the intervention (e.g., a mobile computing platform, a new

workflow process, support from community health workers). Finally, one might assess the ways in which staff satisfaction or dissatisfaction has changed as a result of the intervention. Self-report surveys and interviews might be used to address the issue of satisfaction. Table 3.9 describes research questions, metrics, data, and an approach that can be deployed to measure workforce issues through the dimensions of development and training, deployment, and satisfaction.

Table 3.9. Measurement of Workforce Issues

Research Questions	Core Metrics	Data Sources	Analytic Approach
A. Development and Training			
To what extent do programs provide training to use existing staff versus incorporate new kinds of staff effectively?	Descriptions of innovations, training, evaluations of effectiveness	Interviews with awardee staff and others at institution	Descriptive
Are specialized providers required with training relevant to any of the diseases and systems being targeted?	Descriptions of innovations, training, evaluations of effectiveness	Interviews with awardee staff and others at institution, labor data	Descriptive
What level of investment in training is required to fill these workforce gaps?	Analysis of training and opportunity costs of time and costs of contracted employees	Cost data	Analytic
How effective and efficient are the various training models?	Descriptions of innovations, team staffing, training, effectiveness outcomes variables	Interviews with awardee staff and others at institution, outcome data	Descriptive, analytic
Are providers given feedback on their own performance and relative to others?	Process and outcome patient data aggregated at the provider level	Process and outcome data, focus groups of providers	Descriptive, analytic
Deployment			
To what extent do programs succeed in developing effective work teams that address care needs of the served populations? Are provider-to-provider interactions and discussions more frequent and effective?	Descriptions of innovations, team staffing, training, effectiveness outcomes variables	Interviews with awardee staff and others at institution, self-reports from staff surveys	Descriptive
What is the most effective way to carry out the intervention with patients: to work with patients one-on-one (and in what settings) versus in groups?	Patient contacts must be measured carefully if there are different degrees or levels of "contact" for the intervention	Interviews with awardee staff and others at institution, outcome data	Descriptive, analytic

Research Questions	Core Metrics	Data Sources	Analytic Approach
What are the best ways to contact patients (both from the patient and the provider point of view)?	Descriptions of innovations, team staffing, training, effectiveness outcomes variables	Interviews with awardee staff and others at institution, as well as patients	Descriptive, analytic
Are patients themselves trained on new behavior or interactions with information technology? How do the workers follow up to ensure that the training sticks with the patients (long-term adherence)?	Descriptions of innovations, team staffing, training, effectiveness outcomes variables	Interviews with awardee staff and others at institution, as well as patients	Need longitudinal data over time
Is it more effective to hire new workers or contract for a portion of the time of existing workers in other organizations (or freelance)?	Descriptions of innovations, team staffing, training, effectiveness outcome, and labor cost data	Interviews with awardee staff and others at institution, practice cost data	Descriptive, with cost and accounting methods
Are providers able to work at the "top of their license"?	Provider responses to questions about the intervention and implementation	Interviews with awardee staff and others at institution	Descriptive
Satisfaction			
How has the innovation changed the incidence of burnout among staff? How has the innovation changed incidence of stress among staff?	Staff reports of burnout Staff reports of stress	Assessments of stress or burnout • Surveys • Qualitative interviews with key staff	Parameter estimate, descriptive
What are current rates of staff intent to leave the current practice? How have rates of staff retention and turnover changed over the course of the innovation?	Staff-reported intent to leave Staff turnover and retention	Self-report Administrative and human resources data • Historical data • Current data	Parameter estimate, descriptive Interrupted time series

Research Questions	Core Metrics	Data Sources	Analytic Approach
To what extent are different kinds and levels of staff satisfied or dissatisfied with the care they are able to provide? To what extent are different kinds and levels of staff satisfied with their working conditions? This would include such factors as satisfaction with colleagues, other staff, income, organizational policies, etc. To what extent do different kinds and levels of staff report satisfaction or dissatisfaction with specific components of the intervention? This would include components introduced as part of the intervention (e.g., a mobile computing platform, a new workflow process, support from community health workers). How has staff satisfaction or dissatisfaction changed as a result of the intervention?	Measures of staff satisfaction with care Measures of staff satisfaction with working conditions Measures of staff satisfaction with components of the intervention Changes in satisfaction over the course of the intervention	Self-report from staff • Surveys • Qualitative interviews with staff and providers • Qualitative interviews with key program staff	Parameter estimate, descriptive Descriptive
If the innovation is limited to a subgroup of staff and providers within an organization, what are the unintended consequences and spillover effects on the satisfaction of staff and providers *not* involved in the intervention?	Measures of components of satisfaction (stress and burnout, overall satisfaction, satisfaction with job components) among staff not directly involved in the intervention	Self-report from staff • Surveys • Qualitative interviews with staff and providers	Parameter estimate, descriptive

Impact on Priority Populations

What Are Priority Populations?

Priority populations may include those with certain medical conditions, such as the chronically ill, pregnant women, persons with behavioral health needs, individuals with disabilities, and people living with HIV. Nonmedical priority populations might include homeless individuals, immigrants and refugees, rural populations, ethnic and racial minorities,

non-English-speaking individuals, and underserved groups. Evaluating the impact of HCIA interventions on priority populations means understanding the potential impact of the intervention on these populations, including the impact on clinical outcomes and cost.

Why Is It Important to Evaluate Priority Populations?

A key issue emerging from the HCIA pilots will be decisions about which programs should be replicated and expanded. To make this decision, it will be important to understand the actual and projected impacts of the intervention(s) on priority populations, especially those that are underserved by the health care system.

How Might We Measure Priority Populations?

Two aspects of measuring intervention impact for priority groups are important:

- To what extent were health outcomes, quality, and costs different for priority groups?
- Would outcomes, quality, and cost savings be different for priority groups if the intervention were brought to full scale?

The first, retrospective question requires applying priority group assessments to outcomes, quality, and cost analyses. The second, forward-looking question requires modeling outcomes, quality, and cost outcomes for priority groups across the nation.

The first determination will be to identify priority groups served by the existing awards. In most cases, this should be a relatively straightforward descriptive activity. In some instances, (e.g., for homeless, immigrant, and refugee populations), clear indicators may be less available. The second determination will be an estimate of the total priority population(s) that could be reached if the intervention were rolled out on a national scale. To project the impact of program scale-up, judgments would need to be made about how much of the total population of interest could feasibly be reached in a specific time frame. Table 3.10 describes research questions, metrics, data, and approaches that can be deployed to measure priority populations.

Table 3.10. Measurement of Priority Populations

Research Questions	Core Metrics	Data Sources	Analytic Approach
Medical Priority Groups			
To what extent do the awardee interventions include patients from priority populations?	List of program target populations, program descriptions	Program proposals, awardee administrative reports	Descriptive
To what extent do the awardee interventions address meeting the needs of priority populations as a primary focus?	List of program target populations, program descriptions	Program proposals, awardee administrative reports	Descriptive
To what extent do the awardee interventions focus on addressing the needs of priority populations (e.g., functional limitations that would impact ability to manage conditions)?	List of program target populations, program descriptions	Program proposals, awardee administrative reports	Descriptive
Nonmedical Priority Groups			
To what extent do the awardees address nonmedical priority groups and underserved populations?	List of program target populations, program descriptions	Program proposals, awardee administrative reports	Descriptive
Were awardees able to increase access to care for nonmedical priority groups and underserved populations, and how? In what types of care settings?	Patient characteristics, access to care, program descriptions	Administrative claims, surveys—patients, awardee administrative reports, awardee databases	Analytic, pre-post comparison, difference-in-differences
Are there key underserved populations that were not included in the awardees' patient populations?	List of program target populations, program descriptions	Program proposals, awardee administrative reports	Descriptive

The second determination will be to estimate the impact of the intervention on priority populations. Priority groups may be affected by interventions through a variety of mechanisms, but the outcomes of interest would be observable in differential quality and costs of care across patient groups. The evaluation should also look for both positive and negative changes in priority group outcomes. For example, an intervention could reduce access to care for a minority group but demonstrate increased cost savings for diabetes care overall.

Potential for impact on health outcomes and quality. This part of the evaluation would focus on whether health outcomes and quality were correlated with priority group status. A number of metrics might be used to measure outcomes for priority groups. These include patient characteristics, mortality, morbidity, functional health status, HRQoL, technical quality, rating of providers, rating of provider communication, access to care, care coordination, courtesy and

helpfulness of providers, cultural competency, self-management education, and rating of experience with telehealth equipment and process.

Potential for cost reductions and savings. Cost savings are a major driver for HCIA, and it will be crucial to understand how cost impacts and population size may interact to produce potential savings. For example, small cost reductions that can be applied to large populations may lead to large cost savings systemwide, while large savings that are attributable to very small populations may have limited cost effects systemwide. Table 3.11 describes research questions, metrics, data, and an approach that can be deployed to measure the potential for cost reductions and savings among priority groups.

Table 3.11. Measurement of Impact on Priority Groups

Research Questions	Core Metrics	Data Sources	Analytic Approach
Potential for Impact on Health and Quality Outcomes for Priority Groups			
What are the estimated health and quality outcomes among priority groups?	Patient characteristics Health outcomes (mortality, morbidity, functional health status, HRQoL) Quality outcomes (e.g., rating of providers, rating of provider communication, access to care, care coordination)	Administrative claims, surveys—patients, awardee administrative reports, awardee databases, focus groups of target populations	Comparison across groups DD (pre-post with matched controls if available) Time series (if matched controls not available)
Potential for Cost Reductions and Savings			
What are the estimated cost savings, if any, among priority groups?	Patient characteristics Outcomes from cost assessment	Claims and awardee datasets	Comparison across groups DD (pre-post with matched controls if available) Time series (if matched controls not available)

Context

What Is Context?

Context refers broadly to the environment in which an innovation occurs and, more specifically, to the factors that can help facilitate or impede an intervention's success. A recent issue of *New Directions for Evaluation*, a publication of the American Evaluation Association, defines context as

> [t]he combination of factors (including culture) accompanying the implementation and evaluation of a project that might influence its results, including geographical location, timing, political and social climate, economic conditions, and other things going on at the same time as the project (Fitzpatrick, 2012).

Similarly, a recent study of the role of context in patient safety interventions defines context as "everything else that is not the intervention, but especially those aspects which may influence the intervention implementation and outcomes" (Øvretveit et al., 2012). Context includes endogenous factors, such as leadership, team science, and organizational features, and exogenous factors, such as the policy and political environment in which an intervention is implemented.

Why Is It Important to Measure Context for the HCIA Project?

The evaluation of HCIA programs is concerned with the aspects of context that might influence implementation and outcomes, as well as the ability to sustain and diffuse programs. Key questions focus on the contextual factors that are needed to support a particular intervention: Were there unique characteristics of the awardee organization, market, approaches, or patient populations that affected the implementation and success of the innovation? Was there a clearly designated champion or leader to oversee implementation? These contextual factors will have an impact on program sustainability and the ultimate suitability of the HCIA programs for widespread dissemination and diffusion.

One can imagine strong implementation of a sound innovation, but in a context that is not hospitable to the innovation. For this reason, it is important to understand the contextual factors that cause serious problems or pose barriers to an intervention's success. An understanding of such factors can help to assess the conditions that facilitate or impede an intervention's success and to identify those that are important if the intervention is to be replicated on a large scale. A prominent example was Governor Pete Wilson's efforts to introduce K–12 class-size reductions in California. While there was strong experimental evidence from Tennessee on the gains associated with such reductions, the fact that the approach was implemented in a state with a severe teacher shortage meant that the reform lowered overall teacher quality, leading to weak achievement score gains (Bohrnstedt and Stecher, 2000).

How Might Context Be Measured?

Key dimensions of context include endogenous factors (i.e., awardee characteristics, programmatic changes, leadership, team science, organizational issues) and exogenous factors, such as the policy and political environment. The relevant aspects of context will vary across interventions. We propose to assess context in terms of "fit" or "congruence" between two key elements: (1) the demands and requirements of the innovation and (2) the ability of the organization to meet those demands, given other operational considerations. Thus, the evaluation of context would consider the extent to which leadership, team characteristics, and organizational characteristics supported or conflicted with implementation of the intervention. Also important is the extent to which senior management provided the resources (e.g., staffing, time, funding) needed to implement the innovation. Finally, the evaluation should consider the extent to which the organization's culture supported or conflicted with implementation. Table 3.12 describes research questions, metrics, data, and an approach that can be deployed to measure context.

Table 3.12. Measurement of Context

Research Questions	Core Metrics	Data Sources	Analytic Approach
Endogenous Factors: Leadership			
Was there a clearly designated champion, leader, or point person(s) to oversee implementation?	Description of champion or leader role	Interviews with program staff and senior management, communication materials used by program	Descriptive
To what extent were "point-of-service" providers and patients involved in planning and implementing the innovation? How was the need for the innovation communicated to them?	Description of involvement by providers and patients	Interviews with program staff and senior management, communication materials used by program	Descriptive
To what extent did senior management in the organization provide resources (e.g., staffing, time, funding) needed to implement the innovation?	Description of senior management support	Interviews with program staff and senior management, communication materials used by program	Descriptive

Research Questions	Core Metrics	Data Sources	Analytic Approach
To what extent did implementation of the innovation involve coordination with outside stakeholders (e.g., units and organizations)?	Description of coordination with outside stakeholders	Interviews with program staff and senior management, communication materials used by program	Descriptive
Endogenous Factors: Team Characteristics			
What were the key characteristics of the awardee team that would affect implementation of the innovation?	Description of awardee team structure	Interviews with program staff and senior management, communication materials used by program	Descriptive
Endogenous Factors: Organizational Features			
What were the unique characteristics of the awardee that affected the implementation and success of the innovation?	List of key characteristics of awardee, including • organizational capacity • organizational culture • market • approaches • patient populations	Surveys—staff, surveys—clinicians, structured interviews with staff, structured interviews with clinicians	Qualitative analysis
What key assumptions are required concerning the host organizations' capacities?	List of key assumptions	Review of program documents, literature on similar innovations, and interviews with program developers	Descriptive
To what extent did organizational features support or conflict with implementation?	List of key organizational features	Surveys—staff, surveys—clinicians, structured interviews with staff, structured interviews with clinicians	Qualitative analysis
Endogenous Factors: Stakeholder Engagement			
To what extent did stakeholder engagement affect the relevance, transparency, or adoption of the innovation?	Stakeholders engaged Engagement activities Engagement modes and methods	Written reports and interviews with staff	Qualitative analysis
Exogenous Factors: Policy and Political Environment			
To what extent did the policy and political environment support or conflict with implementation?	List of key policy and political factors	Surveys—staff, surveys—clinicians, structured interviews with staff, structured interviews with clinicians	Qualitative analysis

Chapter 4: Summary Evaluation and Decision Strategy

In addition to the evaluations conducted at the levels of individual awardees and awardee groups, we also see a role for summary evaluation strategies that would include large numbers of awardees or other kinds of awardee groups. For instance, one such group might be awardees that include Medicare recipients as their primary target group. The primary objective of the summary evaluation is to compare and synthesize findings from evaluations conducted at the awardee and grouping levels and from pooled analyses, in order to assist in identifying those interventions that can be implemented more broadly, those that need testing in other settings, and those that may be deferred from current consideration for further investment. This chapter describes options for conducting summary evaluations.

Evaluations of individual awardees and awardee groups, described in previous chapters, can account for unique features of individual innovations and innovation types in ways that a summary evaluation cannot. The benefits of a summary evaluation are the potential to compare, synthesize, and interpret the variety of evaluations that are conducted on individual innovations and smaller groups of awardees. Comparison and synthesis can provide further insight on innovations that are effective at controlling or reducing costs and those that are effective at maintaining or improving health outcomes and quality of care.

There are several secondary objectives of the summary evaluation, and these are to understand (1) how effective innovations can be scaled up to other populations and under what circumstances; (2) what changes in regulations, reimbursement structure, and other policies may be needed to ensure the sustainability of effective innovations; and (3) how less-effective innovations can be tested further, why their outcomes are lacking, and how their outcomes might be improved.

Challenges

There are several challenges associated with conducting a summary evaluation. The first of these has to do with the heterogeneity of awardee activities. Each awardee has proposed and is carrying out multiple, overlapping changes in its health care systems (Table 4.1). Thus, every awardee's innovation is not a single innovation at all, but rather a collection of innovations. Furthermore, the terms used to describe individual innovation types may not be consistent across awardees—e.g., care coordination in one place may not mean the same thing in another place. The heterogeneity of combinations and approaches in awardees' programs presents a significant challenge for the summary evaluation.

Table 4.1. List of Awardee Interventions

App Num	Awardee Name	Care Coord	Care Mgmt	Medical Home	Patient Nav	Home Care	Workflow or Care Process Redesign	HIT	Tele-med	Provider Educ	Provider Payment Reform	Patient Decision Support	Provider Guide-line Adherence	Provider Decision Support	Patient/Prov Shared Decision Making
MANAGEMENT OF MEDICALLY FRAGILE PATIENTS IN THE COMMUNITY **1. Disease/Condition-Specific Targeting**															
Diabetes															
1439	Duke University	Yes	Yes	Yes	–	Yes	Yes	Yes	Yes	–	–	Yes	Yes	Yes	–
2812	FirstVitals Health and Wellness Inc.	Yes	Yes	–	–	–	–	Yes	Yes	–	–	Yes	–	–	–
1487	Joslin Diabetes Center, Inc.	–	Yes	–	–	–	–	–	–	Yes	–	Yes	–	–	–
Cancer															
200	Innovative Oncology Business Solutions, Inc.	–	Yes	Yes	Yes	–	–	Yes	–	–	Yes	Yes	Yes	–	–
1705	The Rector and Visitors of the University of Virginia	–	Yes	–	–	–	Yes	Yes	–	Yes	–	Yes	Yes	Yes	Yes
1275	The Trustees of the University of Pennsylvania	Yes	Yes	–	Yes	Yes	Yes	–	–	Yes	–	–	–	–	Yes
1553	University of Alabama at Birmingham	Yes	Yes	–	Yes	–	Yes	–	Yes	Yes	–	–	Yes	–	–
Childhood Asthma															
1409	Alfred I. duPont Hospital for Child NCC-W of the Nemours Foundation	–	Yes	Yes	–	–	–	Yes	–	Yes	–	Yes	Yes	–	–
1930	Health Resources in Action, Inc.	Yes	Yes	–	–	Yes	–	Yes	–	Yes	Yes	Yes	–	–	–
2075	Le Bonheur Community Health and Well-Being	Yes	Yes	–	Yes	Yes	–	Yes	–	Yes	–	Yes	–	Yes	Yes
Other															
1622	Christiana Care Health Services, Inc.	Yes	Yes	–	–	Yes	–	Yes	Yes	–	–	Yes	Yes	Yes	–
1446	Mountain Area Health Education Center, Inc.	Yes	Yes	Yes	–	–	–	Yes	Yes	Yes	–	Yes	Yes	–	Yes

App Num	Awardee Name	Care Coord	Care Mgmt	Medical Home	Patient Nav	Home Care	Workflow or Care Process Redesign	HIT	Tele-med	Provid er Educ	Provider Payment Reform	Patient Decision Support	Provi-der Guide-line Adhe-rence	Provi-der Deci-sion Sup-port	Patient/Prov Shared Decision Making
2024	Ochsner Clinic Foundation	Yes	–	–	–	Yes	–	Yes	Yes	Yes	–	Yes	Yes	–	–
539	Regents of the University of California, Los Angeles	Yes	Yes	–	Yes	Yes	Yes	Yes	–	Yes	–	–	–	Yes	–
960	The George Washington University	–	Yes	–	–	Yes	Yes	–	Yes	Yes	–	–	–	–	–
874	Trustees of Indiana University	Yes	Yes	–	–	Yes	Yes	Yes	–	Yes	–	–	Yes	Yes	–
653	Upper San Juan Health Service District	–	–	–	–	Yes	Yes	–	Yes	Yes	Yes	Yes	–	–	–
527	Vanderbilt University Medical Center (My Health Team)	Yes	Yes	Yes	–	–	–	Yes	Yes	Yes	–	Yes	–	–	–
2. Complex/High-Risk Patient Targeting															
1078	Beth Israel Deaconess Medical Center	Yes	Yes	–	–	–	Yes	–	–	–	–	Yes	Yes	–	–
592	California Long-Term Care Education Center (SEIU-ULTCW)	–	–	–	Yes	Yes	–	–	–	Yes	–	–	–	–	–
1680	Courage Center	Yes	–	–	–	Yes	Yes	–	Yes	Yes	Yes	Yes	–	Yes	–
2201	Developmental Disabilities Health Services PA	Yes	Yes	Yes	–	–	–	Yes	–	Yes	Yes	–	–	–	–
247	Johns Hopkins University School of Nursing	Yes	Yes	–	–	Yes	–	Yes	–	Yes	–	Yes	–	–	–
2241	LifeLong Medical Care	Yes	Yes	–	–	–	Yes	Yes	–	Yes	–	Yes	–	–	–
1273	North Carolina Community Care Networks, Inc.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	–	Yes	Yes	Yes	Yes	Yes	–
1967	Northland Healthcare Alliance	Yes	Yes	–	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	–	–	–
757	Palliative Care Consultants of Santa Barbara	Yes	–	–	–	Yes	Yes	–	–	Yes	–	–	–	–	–
910	Pittsburgh Regional Health Initiative	Yes	Yes	Yes	–	–	Yes	–	–	Yes	–	Yes	Yes	–	–
579	Providence Portland Medical Center	Yes	Yes	–	Yes	Yes	Yes	Yes	–	Yes	–	–	–	Yes	–
1221	South Carolina Research Foundation	Yes	–	–	–	Yes	–	–	–	Yes	–	Yes	–	–	Yes
2302	St. Francis Healthcare Foundation of Hawaii	Yes	Yes	–	–	–	Yes	Yes	Yes	Yes	–	–	–	–	–

App Num	Awardee Name	Care Coord	Care Mgmt	Medical Home	Patient Nav	Home Care	Workflow or Care Process Redesign	HIT	Tele-med	Provid er Educ	Provider Payment Reform	Patient Decision Support	Provi-der Guide-line Adhe-rence	Provi-der Deci-sion Sup-port	Patient/Prov Shared Decision Making
1026	Suttercare Corporation	Yes	–	–	–	Yes	Yes	–	–	Yes	–	Yes	–	–	Yes
2237	The Johns Hopkins University	Yes	Yes	–	Yes	Yes	–	Yes	–	Yes	–	Yes	–	Yes	–
2041	The University of Texas Health Science Center at Houston	Yes	–	Yes	–	–	–	–	–	Yes	Yes	–	–	–	–
1600	University Emergency Medical Services, Inc.	Yes	–	–	Yes	Yes	Yes	–	–	–	–	–	–	–	Yes
815	University of Arkansas for Medical Sciences	–	–	–	–	Yes	–	–	–	Yes	–	–	–	–	–
1155	University of Iowa	Yes	–	–	–	–	Yes	Yes	Yes	Yes	–	–	–	–	Yes
403	University of New Mexico Health Sciences Center	Yes	Yes	–	–	Yes	Yes	Yes	Yes	Yes	Yes	–	–	–	–
1918	University of North Texas Health Science Center	Yes	Yes	–	–	Yes	Yes	Yes	–	Yes	–	–	–	–	–
1154	University of Rhode Island	Yes	Yes	Yes	–	–	–	Yes	–	Yes	–	Yes	Yes	Yes	–
1067	Vanderbilt University Medical Center	Yes	Yes	–	–	–	Yes	Yes	–	Yes	–	Yes	–	Yes	Yes
3. Behavioral Health															
923	Center for Health Care Services	–	Yes	–	Yes	–	Yes	Yes	–	Yes	Yes	Yes	–	–	Yes
655	Family Service Agency of San Francisco	–	–	–	–	–	Yes	Yes	–	Yes	–	Yes	–	Yes	Yes
1555	Fund for Public Health in New York, Inc.	Yes	Yes	–	Yes	Yes	Yes	–	–	Yes	–	–	–	–	Yes
691	HealthLinkNow Inc.	–	Yes	Yes	–	–	Yes	Yes	Yes	Yes	–	Yes	–	–	–
2120	Institute for Clinical Systems Improvement	Yes	–	–	Yes	–	Yes	Yes	–	Yes	–	–	Yes	Yes	–
2459	Kitsap Mental Health Services	Yes	Yes	–	–	–	Yes	Yes	–	Yes	–	Yes	–	–	–
1086	Maimonides Medical Center	Yes	Yes	Yes	Yes	Yes	Yes	Yes	–	–	Yes	Yes	–	Yes	–
2227	The Feinstein Institute for Medical Research	–	Yes	–	–	Yes	–	Yes	–	Yes	Yes	Yes	–	Yes	Yes
2339	ValueOptions Inc.	–	–	–	Yes	–	Yes	–	–	Yes	Yes	Yes	–	–	–
548	Vinfen Corporation	Yes	Yes	–	–	Yes	Yes	Yes	Yes	Yes	–	–	–	Yes	–

App Num	Awardee Name	Care Coord	Care Mgmt	Medical Home	Patient Nav	Home Care	Workflow or Care Process Redesign	HIT	Tele-med	Provid er Educ	Provider Payment Reform	Patient Decision Support	Provi-der Guide-line Adhe-rence	Provi-der Deci-sion Sup-port	Patient/Prov Shared Decision Making
HOSPITAL SETTING INTERVENTIONS															
4. Condition-Specific Acute Care															
2311	CHRISTUS Health Ark-La-Tex dba CHRISTUS St. Michael Health S	–	–	–	–	–	Yes	Yes	–	Yes	–	–	–	Yes	–
1786	The Methodist Hospital Research Institute	Yes	Yes	–	Yes	Yes	–	Yes	Yes	Yes	–	–	–	Yes	–
453	The Methodist Hospital Research Institute	–	–	–	–	–	Yes	–	–	Yes	–	–	Yes	–	–
1671	Trustees of Dartmouth College—Sepsis	–	–	–	–	–	Yes	Yes	–	Yes	–	Yes	Yes	Yes	Yes
5. Acute Care Management															
1486	Henry Ford Health System	Yes	Yes	–	–	–	Yes	Yes	–	Yes	–	–	Yes	Yes	–
2280	Mount Sinai School of Medicine	Yes	–	–	–	–	Yes	Yes	–	Yes	–	–	–	Yes	–
1832	The University of Chicago	–	–	–	Yes	–	Yes	Yes	–	Yes	–	Yes	–	–	–
6. Improvement in ICU/Remote ICU Monitoring															
1946	Emory University	–	–	–	–	–	Yes	Yes	Yes	Yes	–	–	–	–	–
38	Mayo Clinic	–	Yes	–	–	–	–	Yes	–	Yes	–	–	Yes	Yes	Yes
2343	St. Luke's Regional Medical Center, Ltd.	–	–	–	–	–	Yes	Yes	Yes	Yes	–	–	Yes	Yes	–
COMMUNITY INTERVENTIONS															
7. Community Resource Planning, Prevention, Monitoring															
459	Altarum Institute	Yes	Yes	–	–	–	Yes	Yes	–	Yes	Yes	Yes	Yes	Yes	–
1196	Ben Archer Health Center	–	Yes	Yes	Yes	Yes	–	Yes	–	Yes	–	Yes	–	–	–
2608	Bronx RHIO, Inc.	–	Yes	–	–	–	Yes	Yes	–	Yes	–	–	–	–	–
440	Children's Hospital and Health System, Inc.	Yes	Yes	–	Yes	Yes	Yes	Yes	–	Yes	–	Yes	–	–	–
528	Delta Dental Plan of South Dakota	Yes	Yes	–	Yes	–	Yes	–	–	Yes	–	Yes	–	–	–
2043	Eau Claire Cooperative Health Centers Inc.	–	–	–	Yes	Yes	Yes	–	–	–	–	Yes	–	–	–

App Num	Awardee Name	Care Coord	Care Mgmt	Medical Home	Patient Nav	Home Care	Workflow or Care Process Redesign	HIT	Tele-med	Provid er Educ	Provider Payment Reform	Patient Decision Support	Provi-der Guide-line Adhe-rence	Provi-der Deci-sion Sup-port	Patient/Prov Shared Decision Making
1877	Finity Communications, Inc.	–	Yes	–	Yes	–	Yes	Yes	–	–	–	Yes	–	–	–
525	IHC Health Services, Inc. dba Intermountain Healthcare	–	–	–	–	–	Yes	Yes	–	Yes	Yes	Yes	–	Yes	Yes
2639	Imaging Advantage LLC	–	–	–	–	–	Yes	Yes	Yes	Yes	–	–	–	Yes	–
2768	Mary's Center for Maternal and Child Care, Inc.	–	–	–	Yes	–	Yes	Yes	Yes	–	–	Yes	–	–	–
1568	Michigan Public Health Institute	Yes	Yes	–	Yes	–	Yes	Yes	–	Yes	Yes	Yes	Yes	–	–
2328	Mineral Regional Health Center	Yes	–	–	–	–	Yes	Yes	–	Yes	–	–	–	–	–
59	National Council of Young Men's Christian Associations	–	Yes	–	–	–	–	Yes	–	Yes	–	Yes	–	–	–
2179	Northeastern University	–	–	–	–	–	Yes	–	–	–	–	–	–	–	–
1912	Prosser Public Hospital District	–	–	–	–	Yes	Yes	–	–	Yes	–	–	–	–	–
277	Regional Emergency Medical Services Authority (REMSA)	Yes	Yes	–	–	Yes	Yes	Yes	–	Yes	Yes	–	Yes	–	–
320	South County Community Health Center, Inc.	Yes	Yes	Yes	Yes	Yes	–	Yes	–	Yes	–	Yes	–	–	–
599	Southeast Mental Health Services	–	–	Yes	Yes	–	–	–	–	–	–	Yes	–	–	–
1910	The Asian Americans for Community Involvement of Santa Clara	Yes	Yes	Yes	Yes	–	Yes	–	–	–	–	Yes	–	–	–
885	The Curators of the University of Missouri	–	Yes	–	–	Yes	–	Yes	Yes	Yes	Yes	Yes	Yes	–	–
535	The National Health Care for the Homeless Council	Yes	Yes	Yes	Yes	–	–	–	–	Yes	–	Yes	–	–	–
810	The University of Chicago	Yes	Yes	–	–	–	Yes	–	–	Yes	–	–	–	–	Yes
1631	University of Miami	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	–	–	–
693	Women & Infants Hospital of Rhode Island	–	Yes	Yes	Yes	Yes	–	Yes	–	Yes	–	Yes	–	–	–

8. Primary Care Redesign

App Num	Awardee Name	Care Coord	Care Mgmt	Medical Home	Patient Nav	Home Care	Workflow or Care Process Redesign	HIT	Tele-med	Provid er Educ	Provider Payment Reform	Patient Decision Support	Provi-der Guide-line Adhe-rence	Provi-der Deci-sion Sup-port	Patient/Prov Shared Decision Making
1540	Atlantic General Hospital Corporation	Yes	–	Yes	Yes	Yes	Yes	Yes	–	–	–	–	–	–	–
610	CareFirst, Inc.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	–	–	Yes	–	Yes	Yes	–
110	Cooper University Hospital	Yes	–	–	Yes	–	Yes	–	–	Yes	Yes	–	Yes	Yes	–
2587	Denver Health and Hospital Authority	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	–	–	Yes	–	Yes	–
2531	Finger Lakes Health Systems Agency	Yes	Yes	Yes	Yes	–	–	Yes	–	Yes	Yes	Yes	–	–	–
136	Memorial Hospital of Laramie County dba Cheyenne Regional	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	–	Yes	Yes	–
2100	Pacific Business Group on Health	Yes	Yes	Yes	–	–	–	–	–	Yes	Yes	Yes	–	–	–
67	PeaceHealth Ketchikan Medical Center	Yes	Yes	Yes	–	Yes	Yes	Yes	–	Yes	–	Yes	–	–	Yes
486	Research Institute at Nationwide Childrens Hospital	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	–	Yes	–	–	Yes	–
795	Rutgers, The State University of New Jersey	Yes	Yes	–	Yes	Yes	Yes	Yes	–	–	Yes	Yes	–	–	–
805	Sanford Health	–	Yes	Yes	Yes	–	–	Yes	Yes	–	–	Yes	Yes	Yes	–
2819	San Francisco Community College District—City College of SF	Yes	Yes	–	Yes	Yes	–	Yes	–	Yes	–	Yes	Yes	–	–
2152	TransforMED, LLC	Yes	Yes	Yes	–	–	–	Yes	–	Yes	–	Yes	Yes	Yes	Yes
859	University Hospitals of Cleveland	Yes	–	–	Yes	Yes	Yes	Yes	Yes	Yes	Yes	–	Yes	–	–
9. Pharmacy/Medication Management															
1140	Carilion New River Valley Medical Center	Yes	Yes	–	–	–	Yes	Yes	–	Yes	–	Yes	–	–	–
2648	Pharmacy Society of Wisconsin	Yes	Yes	–	–	–	Yes	Yes	–	Yes	–	Yes	Yes	Yes	–
1129	The Trustees of the University of Pennsylvania	Yes	Yes	–	Yes	–	–	Yes	Yes	–	Yes	Yes	–	Yes	–
1939	University of Southern California	Yes	Yes	–	–	–	Yes	Yes	Yes	Yes	–	Yes	–	–	–
2835	University of Tennessee Health Science Center	Yes	Yes	–	–	Yes	Yes	Yes	–	Yes	–	Yes	–	–	–
10. Shared Decisionmaking															

App Num	Awardee Name	Care Coord	Care Mgmt	Medical Home	Patient Nav	Home Care	Workflow or Care Process Redesign	HIT	Tele-med	Provider Educ	Provider Payment Reform	Patient Decision Support	Provi-der Guide-line Adhe-rence	Provi-der Deci-sion Sup-port	Patient/Prov Shared Decision Making
1921	MedExpert International, Inc.	–	–	–	Yes	–	–	Yes	–	–	–	Yes	–	Yes	Yes
1672	Trustees of Dartmouth College - Patient Engagement	–	–	–	Yes	–	Yes	Yes	–	–	–	Yes	–	Yes	Yes
569	Welvie LLC	–	–	–	Yes	Yes	–	Yes	–	Yes	–	Yes	–	–	Yes

Second, the awardees target a wide range of populations. This includes populations of different age, race and ethnicity, condition, insurance status, and so on. Therefore, care must be exercised in interpreting the potential for scale-up of successful innovations. This challenge, though significant, also presents a major opportunity for a summary evaluation. It is possible that differential effects of similar innovations in different populations may be attributable to variations in the populations, and it may be possible to detect this with a summary evaluation.

Third, awardee innovations and their population impacts will be evaluated in the context of different organizational characteristics. Differences in leadership support, information technology (IT), culture, staffing structure, and so on may be influential on outcomes. Here again, this challenge offers a potential opportunity for the summary evaluation. These differences, if measured carefully, may be detectable only in large-scale comparisons and syntheses of findings from individual awardee and grouping evaluations.

Fourth, and perhaps most challenging for the summary evaluation, individual awardees and evaluators may measure performance differently. In Chapter 3, we presented five major evaluation dimensions on which all HCIA awardees can be evaluated: implementation effectiveness, program effectiveness, workforce issues, impact on priority populations, and context. These dimensions have to be flexibly applied in order to maximize the potential to learn something meaningful from each awardee, innovation, and awardee group. But flexibility in applying these dimensions may result in significant variation in the questions, metrics, data, and approach of the many individual evaluations that are conducted over the course of the program. This means that comparison and synthesis of measurement will be extremely challenging.

Strategy

The summary evaluation strategy has to take account of these challenges. Below we suggest key elements of a strategy that will create opportunities for valid comparison and synthesis of individual awardee and group evaluations.

Coordination of Evaluators

Early coordination of evaluators will be important because it can maximize correspondence and minimize unnecessary variation in the ways that awardee innovations have been assessed, through differences in evaluation questions, metrics, data, or approach. As awardee and group evaluations proceed, coordination will ensure that questions, metrics, data, and approaches are similar enough to produce findings that can be compared and synthesized across the many awardees, awardee groups, and interventions. Coordination would begin with consideration of Table 3.1, to initiate and guide a discussion on the proposed evaluation dimensions. The goal in this first stage would be to use discussion and adjustment to validate what is currently proposed. The process would continue with consideration of Tables 3.2–3.12, to initiate and guide discussions on the research questions, metrics, data, and approaches for evaluation within each of

the grantee groupings. Coordination might be facilitated through scientific guidance, a process for building consensus, or both.

Analysis and Interpretation of Findings

The analysis and interpretation approach we discuss in this chapter is composed of three major components, which can be carried out simultaneously.

Component 1: A Ratings System. An evaluation ratings system may be developed to compare findings from the many qualitative and quantitative measures in grouping, intervention, and program evaluations. This system could be focused on the five major dimensions that are discussed in Chapter 3: implementation effectiveness, program effectiveness, workforce issues, impact on priority populations, and context. The ratings system is designed to summarize findings across evaluation dimensions using different types of data.

Component 2: A Pooled Analysis. Independent assessments of many innovations can be derived from pooled analyses of the populations that are targeted and reached by program awardees, using data from CMS, state, and other administrative or survey sources. This pooled analysis could be focused on the second of five evaluation dimensions that are discussed in Chapter 3: program effectiveness, with focused assessments on the three subdimensions of health, costs, and quality. Pooled analysis is designed to allow for pre- and post-intervention analyses at the organizational, grouping, and innovation levels. Depending on the availability of appropriate and robust data, it may also allow for comparison of intervention populations with matched controls. The pooled analysis is designed to identify key elements of implementation effectiveness by taking advantage of larger sample sizes and comprehensive analytic techniques.

Component 3: A Decision Strategy. The qualitative and quantitative comparisons and syntheses in Component 1 will address opportunities for cross-awardee learning in each of the dimensions presented in Table 3.1. The evaluation approaches that are presented for each question in Tables 3.2 through 3.12 have varying internal and external validity, beyond the individual awardee and group evaluations in which they are first applied. This will influence the strength of evidence assessment that a summary evaluator is able to apply to each evaluation question. The pooled analyses from Component 2 would be focused on program effectiveness and its subdimensions of health, costs, and quality, taking into account opportunities for pooling CMS, state, and other administrative data. A structured decision strategy would use these data to enable systematic consideration of key innovation features and outcomes, in order to develop informed policy. The comparisons and syntheses that arise from pooled analyses have the potential for stronger internal and external validity of findings in the summary evaluation. These pooled analyses can thus be seen as an independent validation of findings from individual awardee, grouping, and Component 1 evaluations.

Summary evaluation may be carried out concurrently with the individual awardee and group evaluations. The main benefits of concurrent work are (1) to ensure coordination among the multiple awardees and evaluators so that they work as much as possible from comparable

evaluation dimensions, questions, metrics, data, and approaches; and (2) to ensure that early findings can be delivered rapidly to CMS, leading to the identification of opportunities for course correction, improvement, and early scale-up if warranted. In order to accomplish this, the evaluators need to be coordinated in their work and have a clear plan for analysis, synthesis, and interpretation of their results.

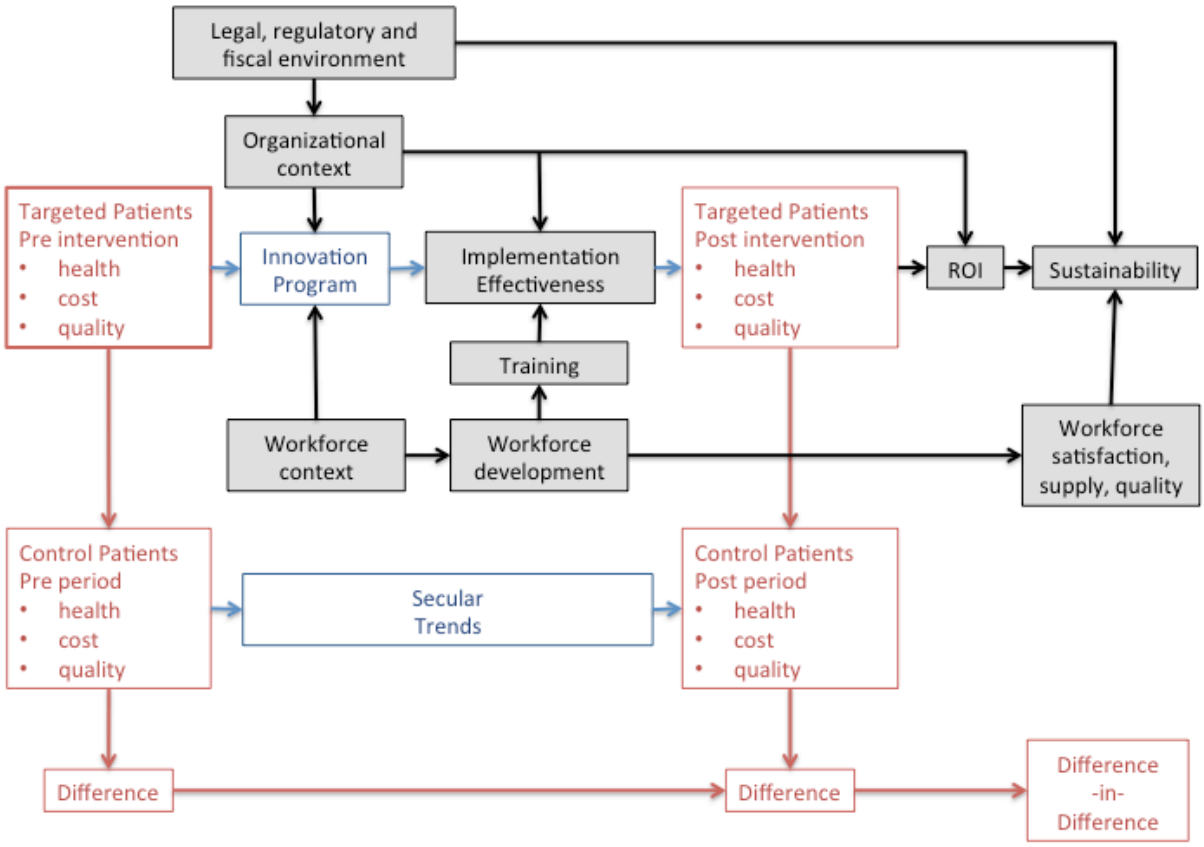
Conceptual Framework and Analysis and Interpretation of Findings

Figure 4.1 expands the conceptual model shown above in Figure S.1 to suggest the presence of a control group that can help to determine the impact of the innovation program on the outcomes of interest. Ideally, this would be a randomized control group selected from among eligible patients at the initiation of the program, but we expect this will be rare among the awardees since most did not undertake this kind of evaluation. However, some may have created or be able to create a post hoc matched control group at their site. In addition, for some analyses it may be possible to create a matched control group from other data—for example, Medicare data that might be used for pooled analyses.

The diagram depicts dimensions pertaining to the three-part aim in red. These dimensions are subject to measurement in individual awardee reports, in pooled analyses, and in evaluator activities. Evaluation dimensions that become inputs into the qualitative case reviews are presented in gray. The case review approach can result in (1) information about why cost reductions were achieved or not, why health was improved or not, and why health care quality was improved or not and (2) information that is relevant for close-out and scale-up. In blue are the interventions and comparison trends that would be compared in a difference-in-difference (DD) analysis.

If it is not possible to develop any control group, then subjects will serve as their own controls. In the event that comparison group(s) are not available or valid, the bottom half of the diagram would not be used and the intervention evaluation would become a cohort design, potentially with the pre-intervention treatment population serving as a control measure and with the post-intervention treatment population serving as the intervention measure.

Figure 4.1. Expanded Conceptual Framework



Component 1: A Ratings System

The purpose of this section is to suggest an evaluation framework that enables direct comparison and synthesis of findings from individual awardee and group evaluations. Given the variation in goals, program structures, populations, interventions, outcomes, and measurement of over 100 HCIA awards, the challenges associated with presenting uniform evaluation standards cannot be overstated. Comparison and synthesis of findings at the program level must cover a heterogeneous group of projects that have been measured on different scales, using different measures. Heterogeneity of programs, populations, and measurement is therefore a chief concern in building this framework. To address these challenges, the comparison and synthesis of findings across multiple awardees and groups will be oriented toward establishing comparable *ratings* and uniform indications of the *strength of evidence* for those ratings.

The next section reviews the challenges and makes recommendations for comparing and synthesizing findings from individual awardee and group evaluations, following the five evaluation dimensions presented in Table 3.1. The final two sections of this first component

discuss a strategy for building ratings and strength of evidence indicators from those comparisons and syntheses.

Evaluation Dimensions

The five evaluation dimensions that were presented in Table 3.1 are program implementation, program effectiveness, workforce issues, priority populations, and context. In this section, we present strategies for comparing and synthesizing findings across awardees and groups within each of these dimensions.

1. Program Implementation

The subdimensions in program implementation—program drivers, intervention, and reach—call primarily for descriptive measures that will vary from awardee to awardee and will be identified in quarterly program reports. These measures will help with decisions about the importance of program implementation factors associated with successful innovations.

Program drivers. What drives a program toward its objectives can be measured through the subdimensions of a program's theory of change and theory of action. Awardees identified drivers in the earliest stages of their program implementation, providing CMS and evaluators with a roadmap that could be used to support evaluation activities. This qualitative information should be extracted from program reports using standardized review procedures, such as review with random quality checks, double review with adjudication, or natural-language processing with secondary review.

Intervention. The nature of a program intervention can be measured through the subdimensions of the intervention's components, dosage, and fidelity and through the awardee's self-monitoring activities. Earlier in this chapter we discussed the challenge of comparing and synthesizing interventions with widely different characteristics, even within interventions that are described by the same name. A review of the components, dosage, and fidelity of interventions can help to determine the comparability of interventions. Information about the intervention components and fidelity is qualitative in nature. This qualitative information can be extracted from program reports using standardized review procedures, such as review with random quality checks, double review with adjudication, or natural-language processing with secondary review. Because of the complexity and heterogeneity of program interventions, natural-language processing may be the most useful strategy for detecting the characteristics of interventions from awardee program and self-monitoring reports. Measures of dosage are typically quantitative but may be presented within individual awardee reports on different scales. Comparison and synthesis of dosage within comparable intervention components is a straightforward task if the scales can be coordinated at the earliest stages of program implementation and evaluation. The goal here is to classify how easy or difficult it was to implement the intervention with fidelity and at the appropriate dosage and what kinds of structures are required to implement the intervention successfully.

Reach. The reach of a program can be measured through the intervention's population coverage and timeliness, as well as through its secondary use of tools. The population coverage of an intervention is a key quantitative measure that is critical to the task of evaluating the potential for scale-up of successful innovations. By comparing the intended and actual reach of a population intervention in different program settings and contexts, evaluators may be able to project upper and lower bounds of an innovation's potential to reach those same populations after nationwide scale-up. The timeliness of the intervention is a fairly simple measure, if applied uniformly as the proportion of a target population that is reached within a specified period of time. This can help to interpret the potential of innovations to reach their intended populations within a time frame that justifies the program investment. The secondary use of tools is a way to assess whether an innovation has substantial potential to improve outcomes and reduce costs to populations that were not initially part of the innovation. This is qualitative information and can be understood as an assessment of the potential for spillover effects, which may be reviewed quantitatively as part of *program effectiveness*, as discussed in the next section. Again, the purpose of a summary evaluation of program reach is to determine how easy or difficult it was to achieve the desired level of coverage in a timely way, and what elements were required to make this possible.

2. Program Effectiveness

Health. Health can be assessed across two distinct evaluation dimensions: health outcomes and HRQoL, as shown in Table 3.1. Both can be used to compare or synthesize findings across awardees and groups. For example, the all-cause mortality rate is a common quantitative measure that is important in most awardee and group assessments. However, even quantitative measures like this may vary significantly across awardee types: Awardees focusing on inpatient care might be interested in inpatient mortality, while those focusing on chronic care management might be interested in annual mortality. Therefore, although mortality rate is a critical metric for both groups, a mortality rate of 10 percent would have a very different meaning for each group.

HRQoL is a dimension for which there are many quantitative and validated measures available. For example, if the SF-36 or SF-12 could be fielded in all awardee groups, that would result in meaningful measures for comparison. However, the comparison of results across groups is complicated. Because each program focuses on a different condition, even a measure that has been shown to be robust (such as the SF-36) may still be differentially sensitive for different conditions and not sufficiently sensitive to detect real improvements in some conditions. Therefore, a disease-specific quality of life measure should also be considered for some sets of awardees.

Costs. Lowering costs through innovation is one of the three primary aims of the HCIA interventions. Calculation of program costs, total costs, and cost savings was addressed in Chapter 3. Total cost calculations include two components: program and utilization costs. Program and total costs are expressed as per beneficiary (or member) per month (or other time

period). Cost savings are the difference of total costs among patients who are enrolled in the intervention and total costs of those who are not. The sustainability of total costs and the cost savings of an intervention can be assessed from the perspective of the payer and the provider, respectively. These assessments have different implications for the decisions that will be made about further investments, further studies, or deferral (Table 4.2).

The standardization of these estimates to a per-beneficiary, per-month estimate is needed to compare and synthesize total costs and cost savings across awardees, groups, and interventions. These numbers, if calculated using similar standards by awardees and evaluators, can be compared directly. They can be synthesized at any level by using the patient (or member) as a denominator. The strength of evidence about costs and savings may be assessed by reviewing the standards used for their measurement.

Table 4.2. Implications of Savings and Sustainability for Policy and Practice

	Provider Perspective: Financially Sustainable	**Provider Perspective: Not Financially Sustainable**
Payer Perspective: Cost-Saving	**Outcome A:** CMS and providers partner to facilitate widespread adoption.	**Outcome B:** CMS modifies regulatory environment and payment policy to create financial incentive for adoption and sustainability.
Payer Perspective: Not Cost-Saving	**Outcome C:** Providers may not implement without regulatory or payment policy changes, unless barriers to cost savings are discovered and removed or facilitators of cost savings are discovered and implemented.	**Outcome D:** Neither stakeholder pursues model without discovering and removing barriers to cost savings and financial sustainability or discovering and implementing facilitators to cost savings and financial sustainability.

Quality. Quality of care can be assessed across six subdimensions: safety, clinical effectiveness, patient experience, timeliness, efficiency, and care coordination, as shown in Table 3.1. The evaluation questions, metrics, data, and approach for individual awardee and grouping evaluations are described in Tables 3.7.1–3.7.6. As with health outcomes, some of these dimensions are relevant for all awardees. For example, awardees may be assessed on the effectiveness of care they deliver to their patients. However, there is substantial variation in the specific measures that can be used to assess clinical effectiveness, and Table 3.7.2 describes this variation.

Awardees focusing on inpatient care will be assessed using a different set of quality measures than those used for awardees focusing on care delivered in the community. Even among programs that deliver care in the community, awardees may focus on particular conditions, such

as diabetes, behavioral health, or cancer care, each of which has a different set of quality measures. Because of this variation, a systematic synthesis for assessing the impact of the awardees and groupings on quality is challenging and complex.

However, most awardees address aspects of quality that have been studied extensively and for which there are established and validated treatment guidelines. These can be reviewed to determine which indicators are most clinically significant and what levels of change might be expected over what period of time in similar interventions. Identification of relevant indicators and treatment guidelines can be used to measure whether changes in quality due to awardee programs are less or greater than expected. This approach, while challenging, offers the opportunity to view different intervention results on similar scales.

3. Workforce Issues

Evaluating the workforce issues raised and discussed in Chapter 3 would involve using a mixture of qualitative and quantitative methods, as described below.

Development, deployment, and training. As awardees report on and describe their activities toward finding, training, and retaining the health care workforce needed for their interventions, the overall evaluation can synthesize these findings using some of the qualitative and quantitative methods noted above. Reports concerning the relative ease or difficulty in recruiting and training providers would be best evaluated using qualitative scales or language-processing methods. Such data as training investment per FTE or turnover or retention rates could be standardized, where the denominator is the number of provider FTEs specifically associated with the intervention. However, the degree of standardization would be limited, given the heterogeneity of interventions and difficulty in associating portions of time of new and existing providers with the intervention itself.

The way in which workforces are deployed is more of a program feature than an outcome to be evaluated normatively. However, depending on the extent to which classification is useful in analysis and potential scale-up of the intervention, evaluators could seek to classify deployment into broad categories, such as the ways in which they interact with patients, the extent to which they work with teams (versus acting alone), and the extent to which they are working at the top of their license.

Worker satisfaction. Measures of provider satisfaction are more amenable to quantitative assessment than workforce development, deployment, and training. Chapter 3 notes some of the research and scales used particularly for physicians. Similar measures exist for nonphysician providers. A summary evaluation may be able to compress or align these measures on a consistent scale to assess patterns or broad measures of positive or negative satisfaction. Awardees may take different approaches to determining which providers should be used to measure satisfaction changes. In comparing satisfaction across awardees, the evaluation should consider the extent to which providers are either more directly or more tangentially involved in the intervention.

4. Priority Populations

One key evaluation challenge is to assess the impact of awardee interventions on priority populations, which includes understanding (1) disparities in health, cost, and quality outcomes and (2) the impact of cost savings and quality improvements in these populations after scale-up to a national level. The aim of achieving better care, better health, and lower costs applies in general to the entire populations targeted by awardee interventions. This aim also applies to priority and at-risk populations, who may represent a subset of the whole populations targeted by the interventions.

Priority and at-risk populations may include underserved groups, such as minorities and low-income families, women, children, patients in advanced care or with rare diseases, military service members, and several other categories that are described in Chapter 3. Awardee interventions may target some or all of these populations. A key challenge from the outset will be to assess the extent to which priority and at-risk populations are targeted and reached by awardee interventions. To address this challenge, we propose two steps. In the first step, we propose to classify each awardee into one of three groups: (1) those that have no emphasis on priority and at-risk populations, (2) those that include some priority and at-risk populations as a portion of their total patient population, and (3) those that focus exclusively on priority and at-risk populations. This first step yields an output that is helpful to decisionmakers in identifying key projects and possibly key groups of interest. In the second step, we propose to identify and quantify the number of patients from priority and at-risk populations who are targeted and reached by awardee interventions. This second step yields an output that is helpful in projecting the potential effects on priority populations of scale-up of the intervention to a national level. For example, "high utilizers" may be a group where reductions in cost are achieved.

Comparison and synthesis of the impact of awardee interventions on priority and at-risk populations will depend on broadly utilized, nationally recognized measures and those in use across awards. Examples include total cost of care, the Prevention Quality Indicators (PQI), the Health Effectiveness Data and Information Set (HEDIS), and the Consumer Assessment of Healthcare, Providers and Systems (CAHPS).

5. Context

The context in which a program operates can be assessed through the subdimensions of organizational leadership, team characteristics, organizational features, and stakeholder engagement. This qualitative information can be extracted from program reports and interviews with program leaders using standardized review procedures, such as review with random quality checks, double review with adjudication, or natural-language processing with secondary review. Comparison and synthesis of context across awardees, awardee groupings, and interventions may best be achieved if standard frameworks for assessing leadership, organizational features, and stakeholder engagement are agreed upon in advance. Several strong organizational (Helfrich, 2007; Hartnell, 2011) and stakeholder engagement (Concannon, 2012) frameworks are readily

available for selection at the outset of program implementation and evaluation (Concannon et al., 2012; Deverka et al., 2012).

Development of a Common Scale and Ratings System

In this section, we present a common scale and ratings system that draws from the findings on quantitative measures within each of the five evaluation dimensions. Each of these measures could be reinterpreted by the evaluator on a five-point scale centered on 0 (–2, –1, 0, 1, 2) to represent strongly negative to strongly positive change following implementation of the innovation program. A potential scoring system that would summarize results across health, cost, and quality measures for a particular awardee is presented in Table 4.3.

Table 4.3. Scoring Criteria

Score	Interpretation
–2	Consistent significant deterioration in multiple measures of quality with substantial clinical impact
–1	Significant deterioration in some measures of varying clinical impact
0	No change
+1	Significant improvement on some measures of varying clinical impact
+2	Consistent significant improvement on multiple measures with substantial clinical impact

The scores can then be assigned to a vector of three attributes—health, cost, and quality—attributable to each of the awardees, groupings, or innovations of interest. In Table 4.4, Panel A, we show hypothetical measures that contribute to this vector in rows for awardees #1 and #2, using hypothetical scores. These vectors will be taken up again in Component 3 of the summary evaluation, shown in the Panel B.

Table 4.4. The Ratings System

Panel A—A Ratings System (Component 1)			Panel B—A Decision Strategy (Component 3)
Awardee, Grouping, or Innovation Type	**Measure (health, cost, quality, etc.)**	**Common Scale (–2, –1, 0, +1, +2)**	**A. Significant improvement = +1 B. Indeterminate improvement = 0 C. Significant decline = –1**
Awardee 1	Health measure 1 (e.g., all-cause mortality)	0	
	Health measure 2 (e.g., functional status)		
	Cost measure 1 (e.g., program costs)	1	
	Cost measure 2 (e.g., utilization)	-	
	Quality measure 1 (e.g., adverse events)	1	
	Quality measure 2 (e.g., CAHPS)		
Awardee 2	Health measure 1 (e.g., all-cause mortality)	0	
	Health measure 2 (e.g., functional status)		
	Cost measure 1 (e.g., program costs)	–2	
	Cost measure 2 (e.g., utilization)	-	
	Quality measure 1 (e.g., adverse events)	1	
	Quality measure 2 (e.g., CAHPS)		

Development of Uniform Strength of Evidence Indicators

In some circumstances, the data available to individual awardees will support a more detailed evaluation than is possible in pooled analyses. For example, some awardee programs operate in settings in which it will be possible to use quasi-experimental methods to measure program impacts, because they have been identified prospectively and are following over time a population of sufficient size with a well-matched comparison group. In some cases, awardees may have used random sampling as way of enrolling patients and randomization as a means of

assigning participants to intervention and control groups, making it possible to use the most rigorous possible experimental methods to measure program impacts.[1]

However just as the populations, interventions, comparison groups, and *a priori* outcomes of interest vary widely across awardees, so do the quality and rigor of their evaluation designs. Awardees with small numbers of participants will be less able to draw statistically reliable conclusions about program impacts. Without well-constructed comparison groups, evaluators will have difficulty identifying the effects of interventions on awardee populations. The certainty of benefit in awardee programs without well-defined *a priori* outcomes of interest may be difficult to assess.

This lack of consistency in awardee-level data makes rankings and side-by-side comparisons difficult to interpret and potentially misleading. For example, it may not be clear which program is more effective: a small program with a large estimate of cost savings but a large confidence interval surrounding the estimate, or a large program with a small estimate of cost savings and a small confidence interval around the estimate.

To establish some consistency in the estimate of strength of evidence, the summary evaluator should have a rating system in mind, such as the GRADE system (Atkins, 2004) or the AHRQ methods guide on grading strength of evidence (Owens, 2009). This ratings system asks evaluators to assess evidence on the basis of study limitations (risk of bias), consistency, directness, precision, and reporting bias (publication, outcome, and selective analysis reporting bias). Additional domains can include dose-response, observed and unobserved confounders, and strength (magnitude) of association. In these models, studies are typically graded with high, moderate, low, or insufficient strength of evidence.

To assess risk of bias, it will be helpful to refer to a hierarchy of evidence model. Several are available from the evidence-based medicine (EBM) literature (Sackett et al., 1996) that suggest increasing potential bias with a specific order of study designs. There are many EBM hierarchies available. Most agree on the following approximate order of designs:

1. Pooled analyses of awardee randomized controlled trials
2. Individual awardee randomized controlled trials
3. Pooled analyses of awardee quasi-experiments with well-matched comparison group
4. Individual awardee quasi-experiments with well-matched comparison group
5. Pooled analyses of awardee cohort studies without comparison group
6. Individual awardee cohort studies without comparison group
7. Individual or multiple case studies based on qualitative data

This hierarchy can help to address the issue of risk of bias, but the evaluator should consider other external dimensions as well. Issues such as the consistency of intervention, directness in

[1] For example, the University of Chicago, "CommunityRx System: Linking Patients and Community-Based Service," uses random sampling methods (CMS, undated).

the relationship between intervention and outcome (as put forth in the program drivers model and in the judgment of the evaluator), precision of measurement, and potential bias in the reporting of results are all important potential factors. The effectiveness and the external validity of an intervention may also be important to consider. Information about the strength of association between intervention and outcome, the magnitude of effect, the presence of confounders, and the potential for biased measurement and reporting are also important to consider. An awardee whose data address a very narrowly defined population (geographically or in terms of age, gender, or other characteristics) may provide less information overall than one that addresses a broader range of the relevant population of interest.

Component 2: A Pooled Analysis

When the outcomes of interest can be standardized across awardees and comparison groups can be consistently constructed, it can be informative to pool data from multiple awardees to generate rigorous evidence about the impact of interventions, settings, and contexts on outcomes that cannot be obtained readily at the awardee level. Analysis of pooled data can also support evaluation of the comparative effectiveness of different interventions, improve the statistical efficiency of award-level impact estimates, and offer greater flexibility than traditional meta-analysis. More specifically, such a pooled analysis can generate (1) estimates of average impact of intervention type, setting, and implementation context on outcomes; (2) comparisons of the effectiveness of different interventions; and, (3) in some cases, more-reliable estimates of award-level impacts than would be possible with awardee-specific data alone.

Advantages

Awardee-level estimates of program impacts provide information about the effectiveness of individual programs in the specific settings and populations in which they are implemented. Such detailed, program-specific information is of paramount importance to awardee-level program managers who must assess whether CMS-funded programs achieved their intended outcomes and whether they can be sustainable in the absence of CMS funds.

Pooled analysis can supplement awardee-level analysis by compensating for variation in the availability and specificity of data at the awardee level and addressing evaluation questions of specific interest to CMS that cannot be readily informed through awardee-level analysis. In the paragraphs below, we describe how analysis of pooled data at the group and initiative levels can potentially widen the scope of evaluation questions and improve the reliability of inferences about program effectiveness.

Impact of specific interventions, settings, and contexts. Programs funded by CMS employ up to ten distinct intervention types (e.g., care coordination, care management, use of HIT, shared decisionmaking). Data on the independent contribution of these features (and combinations of features) is crucial for informing decisions about how to refine and disseminate

innovative practice models. Examples of evaluation questions about specific intervention features include the following:

- How effective on average were hospital-based interventions in reducing hospital costs?
- How effective on average were interventions that involved care coordination in the community setting in reducing total medical care costs?

Such analyses of the impact of intervention type and models are not possible at the awardee level and may be limited even at the grouping level because of purposeful clustering of similar models within groupings. For instance, are certain interventions more successful in one region of the country than in others? Does the impact of community-based interventions vary by population densities?

Comparative effectiveness of alternative innovation models. While not of specific interest to award-level program managers, it is important to compare the effectiveness of different program models and features when caring for a clinically similar population in similar types of settings. For instance, CMS may be interested in understanding whether post-discharge care-coordination or primary care medical homes have a larger effect on hospital utilization and costs. This type of question cannot be answered at the awardee level but may be answered at the group level, depending on the degree of heterogeneity of programs within groups and the number of participants.

Flexibility. Assuming that treatment effects can be reliably measured at the awardee level, evaluators can use traditional meta-analysis techniques to obtain a summary average effect across all individual effect estimates. While quick to implement, meta-analysis offers very limited capability to examine the effectiveness of specific program features or to compare program impacts for subgroups of studies selected on the basis of program features.

Statistical efficiency. As shown in Figure 4.2 and Table 4.5, many awardees projected in their proposals and work plans to have small numbers of patients involved in their implementations. Pooled analysis can help to address the problem of limited sample at the award level, especially when analyzing the impact of interventions on subgroups of participants. Depending on the magnitude of anticipated impacts and variation in the patient population of interest, large samples (of 5,000 or more) of participants and controls may be required to detect true impacts of awardee programs on outcomes of interest (see Appendix B). This is especially true for utilization and cost outcomes, which are highly variable in the general patient population.

Figure 4.2. Projected Number of Participants by Award and Sample Size Requirements

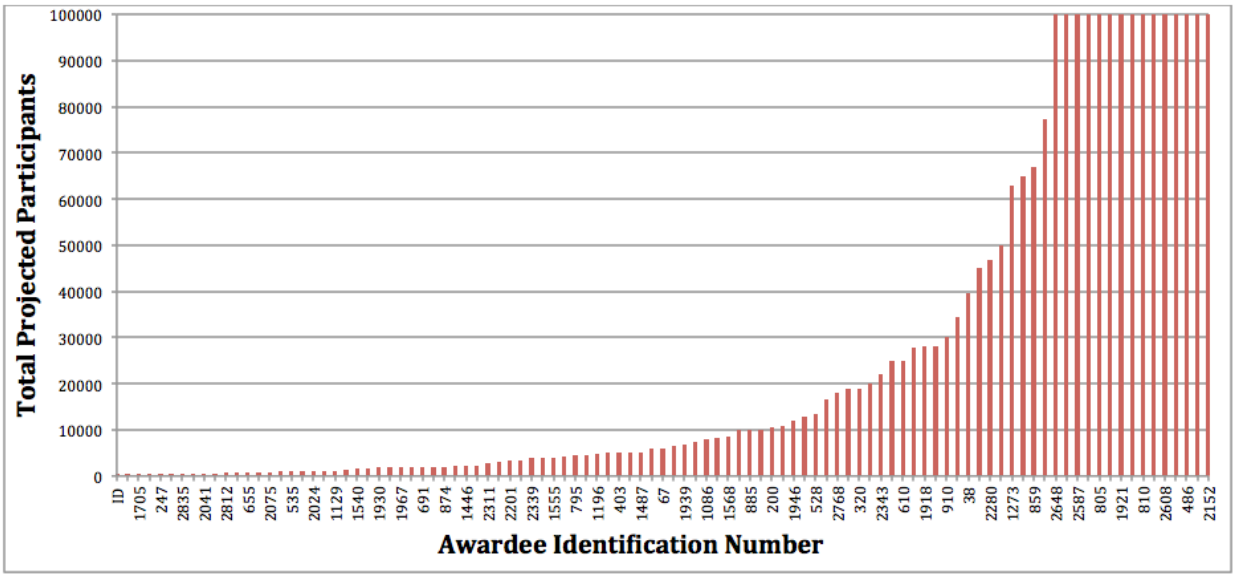


Table 4.5. Participants by Awardee

App Num	Awardee Name	Total Participants	Number of Medicare Beneficiaries	Number of Medicaid Beneficiaries	Medicare	Medicaid	CHIP	Other Payers
MANAGEMENT OF MEDICALLY FRAGILE PATIENTS IN THE COMMUNITY								
1. Disease/Condition-Specific Targeting								
Diabetes								
1439	Duke University	67,080	–	–	Yes	Yes	No	No
2812	FirstVitals Health and Wellness Inc.	600	–	–	No	Yes	No	Yes
1487	Joslin Diabetes Center, Inc.	5,100	–	2,976	Yes	Yes	No	Yes
Cancer								
200	Innovative Oncology Business Solutions, Inc.	19,016	15,961	–	Yes	Yes	No	No
1705	The Rector and Visitors of the University of Virginia	960	768	–	Yes	No	No	Yes
1275	The Trustees of the University of Pennsylvania	2,000	1,000	200	Yes	Yes	No	Yes
1553	University of Alabama at Birmingham	16,553	–	–	Yes	Yes	No	No
Childhood Asthma								
1409	Alfred I. duPont Hospital for Child NCC-W of the Nemours Foundation	4,059	–	4,059	No	Yes	No	No
1930	Health Resources in Action, Inc.	1,462	–	1,039	No	Yes	Yes	Yes
2075	Le Bonheur Community Health and Well-Being	800	–	–	No	Yes	No	No

App Num	Awardee Name	Total Participants	Number of Medicare Beneficiaries	Number of Medicaid Beneficiaries	Medicare	Medicaid	CHIP	Other Payers
Other								
1622	Christiana Care Health Services, Inc.	6,000	3,067	–	Yes	Yes	No	No
1446	Mountain Area Health Education Center, Inc.	2,241	–	–	Yes	Yes	No	No
2024	Ochsner Clinic Foundation	983	–	–	Yes	Yes	Yes	Yes
539	Regents of the University of California, Los Angeles	2,500	–	–	Yes	No	No	No
960	The George Washington University	300	–	–	Yes	Yes	No	No
874	Trustees of Indiana University	2,000	2,000	–	Yes	Yes	No	No
653	Upper San Juan Health Service District	3,443	–	–	Yes	Yes	Yes	No
527	Vanderbilt University Medical Center (My Health Team)	140,000	75,548	–	Yes	No	No	No
2. Complex/High-Risk Patient Targeting								
1078	Beth Israel Deaconess Medical Center	8,047	–	–	Yes	Yes	No	No
592	California Long-Term Care Education Center (SEIU-ULTCW)	18,000	–	–	Yes	Yes	No	No
1680	Courage Center	300	–	–	No	Yes	No	No
2201	Developmental Disabilities Health Services PA	3,072	–	–	Yes	Yes	No	Yes
247	Johns Hopkins University School of Nursing	500	–	–	Yes	Yes	No	No
2241	LifeLong Medical Care	9,750	–	2,500	Yes	Yes	No	No
1273	North Carolina Community Care Networks, Inc.	0	–	–	No	Yes	Yes	No
1967	Northland Healthcare Alliance	1,840	–	–	Yes	Yes	No	Yes
757	Palliative Care Consultants of Santa Barbara	3,200	–	–	Yes	Yes	No	No

App Num	Awardee Name	Total Participants	Number of Medicare Beneficiaries	Number of Medicaid Beneficiaries	Medicare	Medicaid	CHIP	Other Payers
910	Pittsburgh Regional Health Initiative	19,000	19,000	–	Yes	No	No	No
579	Providence Portland Medical Center	39,600	–	–	No	Yes	No	No
1221	South Carolina Research Foundation	1,600	–	–	Yes	Yes	No	No
2302	St. Francis Healthcare Foundation of Hawaii	2,790	–	–	Yes	Yes	No	Yes
1026	Suttercare Corporation	10,738	–	–	Yes	Yes	No	Yes
2237	The Johns Hopkins University	95,000	–	–	Yes	Yes	No	Yes
2041	The University of Texas Health Science Center at Houston	525	–	–	No	Yes	Yes	No
1600	University Emergency Medical Services, Inc.	2,400	–	–	Yes	Yes	No	No
815	University of Arkansas for Medical Sciences	4,200	–	–	Yes	Yes	No	No
1155	University of Iowa	5,000	–	–	Yes	Yes	No	No
403	University of New Mexico Health Sciences Center	5,000	–	5,000	No	Yes	No	No
1918	University of North Texas Health Science Center	27,799	–	–	Yes	Yes	No	No
1154	University of Rhode Island	1,929	–	–	Yes	Yes	No	No
1067	Vanderbilt University Medical Center	27,897	–	–	Yes	No	No	No

3. Behavioral Health

923	Center for Health Care Services	260	–	–	No	Yes	No	No
655	Family Service Agency of San Francisco	765	–	268	No	Yes	No	No
1555	Fund for Public Health in New York, Inc.	3,833	–	–	Yes	Yes	No	No
691	HealthLinkNow Inc.	6,000	–	–	Yes	Yes	Yes	No
2120	Institute for Clinical Systems Improvement	29,332	–	–	Yes	Yes	No	No

App Num	Awardee Name	Total Participants	Number of Medicare Beneficiaries	Number of Medicaid Beneficiaries	Medicare	Medicaid	CHIP	Other Payers
2459	Kitsap Mental Health Services	1,000	–	–	Yes	Yes	Yes	No
1086	Maimonides Medical Center	7,500	–	–	Yes	Yes	No	Yes
2227	The Feinstein Institute for Medical Research	770	–	–	No	Yes	No	No
2339	ValueOptions Inc.	14,994	–	14,994	No	Yes	No	No
548	Vinfen Corporation	1,410	–	–	Yes	Yes	No	No

HOSPITAL SETTING INTERVENTIONS

4. Condition-Specific Acute Care

2311	CHRISTUS Health Ark-La-Tex dba CHRISTUS St. Michael Health S	19,050	–	–	Yes	Yes	No	No
1786	The Methodist Hospital Research Institute	52,450	–	–	Yes	No	No	No
453	The Methodist Hospital Research Institute	106,806	–	–	Yes	Yes	No	Yes
1671	Trustees of Dartmouth College—Sepsis	–	–	–	Yes	No	No	No

5. Acute Care Management

1486	Henry Ford Health System	11,000	–	–	No	No	No	No
2280	Mount Sinai School of Medicine	135,000	135,000	–	Yes	No	No	No
1832	The University of Chicago	2,500	1,500	–	Yes	No	No	No

6. Improvement in ICU/Remote ICU Monitoring

1946	Emory University	10,963	–	–	Yes	Yes	No	No
38	Mayo Clinic	34,322	–	–	Yes	Yes	No	No
2343	St. Luke's Regional Medical Center, Ltd.	16,660	–	–	Yes	Yes	No	No

COMMUNITY INTERVENTIONS

App Num	Awardee Name	Total Participants	Number of Medicare Beneficiaries	Number of Medicaid Beneficiaries	Medicare	Medicaid	CHIP	Other Payers
7. Community Resource Planning, Prevention, Monitoring								
459	Altarum Institute	1,400,000	–	–	No	No	No	No
1196	Ben Archer Health Center	4,656	–	–	Yes	Yes	Yes	No
2608	Bronx RHIO, Inc.	350,000	–	–	Yes	Yes	No	No
440	Children's Hospital and Health System, Inc.	13,600	–	13,600	No	Yes	Yes	No
528	Delta Dental Plan of South Dakota	39,000	–	–	No	Yes	Yes	Yes
2043	Eau Claire Cooperative Health Centers Inc.	4,800	–	–	Yes	Yes	No	Yes
1877	Finity Communications, Inc.	121,450	–	121,450	No	Yes	No	No
525	IHC Health Services, Inc., dba Intermountain Healthcare	919,455	–	–	Yes	Yes	Yes	Yes
2639	Imaging Advantage LLC	28,008,678	–	–	Yes	Yes	No	No
2768	Mary's Center for Maternal and Child Care, Inc.	10,800	–	–	No	Yes	No	No
1568	Michigan Public Health Institute	13,311	–	–	Yes	Yes	No	No
2328	Mineral Regional Health Center	458,324	–	–	Yes	Yes	Yes	No
59	National Council of Young Men's Christian Associations	10,000	10,000	–	Yes	No	No	No
2179	Northeastern University	5,041,667	–	–	Yes	Yes	No	No
1912	Prosser Public Hospital District	800	–	–	Yes	Yes	No	No
277	Regional Emergency Medical Services Authority (REMSA)	8,500	–	–	Yes	Yes	No	No
320	South County Community Health Center, Inc.	19,527	–	–	Yes	Yes	Yes	No
599	Southeast Mental Health Services	4,384	–	–	Yes	Yes	Yes	No

App Num	Awardee Name	Total Participants	Number of Medicare Beneficiaries	Number of Medicaid Beneficiaries	Medicare	Medicaid	CHIP	Other Payers
1910	The Asian Americans for Community Involvement of Santa Clara	12,500	–	–	Yes	Yes	Yes	No
885	The Curators of the University of Missouri	31,318	–	–	Yes	Yes	No	No
535	The National Health Care for the Homeless Council	969	–	–	Yes	Yes	No	No
810	The University of Chicago	211,788	–	–	Yes	Yes	Yes	Yes
1631	University of Miami	36,000	–	–	No	Yes	Yes	No
693	Women & Infants Hospital of Rhode Island	2,400	–	440	No	Yes	No	Yes

8. Primary Care Redesign

1540	Atlantic General Hospital Corporation	1,314	1,314	–	Yes	No	No	No
610	CareFirst, Inc.	87,000	–	–	Yes	Yes	No	No
110	Cooper University Hospital	1,230	–	–	Yes	Yes	No	No
2587	Denver Health and Hospital Authority	147,890	–	–	Yes	Yes	Yes	No
2531	Finger Lakes Health Systems Agency	1,277,681	–	–	Yes	Yes	No	Yes
136	Memorial Hospital of Laramie County dba Cheyenne Regional	265,044	–	–	Yes	Yes	Yes	No
2100	Pacific Business Group on Health	27,000	–	–	Yes	Yes	No	No
67	PeaceHealth Ketchikan Medical Center	7,512	–	–	No	No	No	No
486	Research Institute at Nationwide Childrens Hospital	1,552,155	–	1,552,155	No	Yes	No	No
795	Rutgers, The State University of New Jersey	4,225	–	–	Yes	Yes	No	No
805	Sanford Health	308,138	–	–	Yes	Yes	Yes	No
2819	San Francisco Community College District—City College of SF	5,336	–	–	Yes	Yes	No	No

App Num	Awardee Name	Total Participants	Number of Medicare Beneficiaries	Number of Medicaid Beneficiaries	Medicare	Medicaid	CHIP	Other Payers
2152	TransforMED, LLC	2,262,856	–	–	Yes	Yes	Yes	Yes
859	University Hospitals of Cleveland	196,500	–	133,700	No	Yes	No	Yes
9. Pharmacy/Medication Management								
1140	Carilion New River Valley Medical Center	4,000	–	–	Yes	Yes	Yes	Yes
2648	Pharmacy Society of Wisconsin	77,400	–	–	No	Yes	No	Yes
1129	The Trustees of the University of Pennsylvania	1,038	–	–	Yes	Yes	No	Yes
1939	University of Southern California	40,560	–	–	Yes	Yes	No	No
2835	University of Tennessee Health Science Center	1,175	–	–	Yes	Yes	No	No
10. Shared Decisionmaking								
1921	MedExpert International, Inc.	160,000	–	–	Yes	No	No	No
1672	Trustees of Dartmouth College—Patient Engagement	–	–	–	Yes	Yes	No	No
569	Welvie LLC	160,000	160,000	–	Yes	No	No	No

Data Sources and Measures

Pooled analysis requires measures that are standardized across awardees. As such, this analysis is likely to be feasible only for a subset of program outcomes, participants, and features. Standardized data on utilization- and cost-related outcomes can be obtained from claims data over comparable pre- and post-implementation time periods. Standardized data on health care quality can be obtained from standardized assessments (e.g., CAHPS surveys and HEDIS measures) that are administered uniformly across all awards. Such an approach would exclude uninsured participants and those newly enrolled in their health insurance plans (e.g., those newly eligible for Medicare). Such data may be most readily obtained for Medicare beneficiaries participating in HCIA award programs. While it is technically possible to obtain claims data from commercial insurers and state Medicaid and CHIP programs, arrangements to obtain such data would need to be made on a plan-by-plan basis. Even if restricted to Medicare, estimates of program impacts based on enrollment data reported to CMS in proposal and post-award documents suggest that a pooled analysis that focuses solely on Medicare beneficiaries would include more than 425,000 clinically diverse participants widely distributed over a diverse array of award programs and program features.

Data abstracted from proposal and award documents and coded by RAND, as well as other evaluation-related data, can serve as a potential source of standardized data to inform impacts by program attributes and context (Morganti et al., 2013). To ensure accuracy, however, data may need to be verified and updated by awardees prior to finalizing data analysis, or this information may be available from individual- or group-level evaluations. The pooled analysis will focus on intervention types and their interactions with case mix, setting, staffing, geography, and other covariates. Program attributes of potential evaluation interest include

- Patient characteristics (through in difference-in-difference analysis, these characteristics will be used to identify matched controls)
- Intervention type (see Table 4.1) and any additional intervention features or elements that are distinct from intervention type (e.g., use of nonphysician providers)
- Staffing
- Care setting indicators
- Geographic indicators
- Implementation effectiveness
- Other contextual factors, such as leadership, teamwork, and workforce issues.

Comparison Groups

Administrative claims data can be used to measure standardized utilization and cost outcomes and to construct comparison groups for a pooled analysis in a manner that is consistent across award programs. In the absence of randomization, propensity score matching methods can

be used to identify individuals in claims data who are clinically, sociodemographically, and geographically similar to those participating in HCIA programs. Propensity score methods work by estimating (after applying inclusion and exclusion criteria) the probability that an individual represented in claims data participated in an HCIA award program, based on characteristics of individuals that are observable in claims data. Estimated probabilities are used to create weights that are then used to identify a subgroup of individuals who are most similar to program participants, but who did not participate in an intervention, to serve as a basis of comparison in estimating impacts.

Interpretation of Impact Estimates

Interpretation of estimates is difficult when interventions are not truly the same across awardees, even though they are categorized the same way. For example, in the case of care coordination, some awardees may use a registered nurse coordinator, while others may rely on physicians for this role, and still others may use a licensed practical nurse. Some may use an interactive electronic medical record (iEMR) to reach patients in coordination activities, while others may use the telephone only. Even if a factor is defined in the same way, awardees may have experienced different levels of exposure (i.e., dosages), which may not have been recorded. Some of these differences can be measured and included as additional control variables. Nevertheless, the pooled analysis cannot easily capture all the ways in which interventions vary. The potential statistical approaches we suggest for the pooled analysis all have some mechanisms to allow for deviations of a specific effect with respect to the summary effect. For example, a hierarchical model for pooled analysis can use a random effect or nested fixed effects. It is also well known that various propensity-score approaches have already assumed differential effects at different levels of aggregations, thus inherently allowing for such deviations.

Component 3: A Decision Strategy

Component 3 is designed to knit information from Components 1 and 2 into a decision strategy for meeting the primary objectives of the summary evaluation: *to compare and synthesize findings from evaluations conducted at the awardee and grouping levels, in order to assist in identifying those interventions that can be implemented more broadly, those that need testing in other settings, and those that may be deferred from current consideration for further investment*.

The approach described in this section follows from the objectives of the evaluation. It offers a method to establish a common sense interpretation of comparable measures where they are available, transparent ratings on a common scale where comparable measures are not available, and a case review process that can help CMS and evaluators interpret why a program was effective and how it might best be scaled up. The goal is to summarize results within and across

awardees in as consistent a way as possible to allow a structured and transparent decisionmaking process.

We suggest consideration of a modified multi-criteria analysis (MCA) framework (Department of Communities and Local Governments, 2009). An MCA framework can be used to structure the analytic work and findings from Components 1 and 2 into this decision strategy. An MCA is a suite of decisionmaking tools for use where uncertainty exists, where decisions must be on the basis of multiple evaluation measures, or where measures are not consistent across evaluation subjects. The MCA framework supports objective comparison of interventions on the basis of *a priori* criteria that represent the interventions' summary goals or priorities. A modified MCA approach to address the HCIA Innovations Program would involve the following steps:

Step 1: Establish the decision context. Establishing the decision context in is an important step taken to determine the aims of the MCA process and establish the key decisionmakers. The decision context for evaluation of HCIA awardees is represented by the five dimensions in Table 3.1 and discussed throughout Chapters 3 and 4.

Step 2: Identify value and performance criteria. The second step is to determine the criteria on which the performance of each program will be assessed; the summary evaluation can proceed by listing the measures that apply to each awardee. Emphasis will be placed on measures that are included in the CMS preferred measures list and that have been selected by awardees for self-monitoring. In the evaluation dimensions of program implementation, workforce issues, priority populations, and context, the same measures will apply to each of the awardees, whereas in the dimension of program effectiveness, the measures will be vary across awardees, depending on populations and interventions of interest.

Step 3: Rate performance of each awardee against the criteria. Each awardee's performance will be rated against the measures, and two factors will be considered: size of the effect and strength of evidence for the effect, as discussed in the last two sections of Component 1. Scores on each of these factors need to be developed and then applied to each awardee for each measure.

Step 4: Create summary scores. All the scores for each awardee can then be combined to provide summary scores for performance and strength of evidence on each of the five evaluation dimensions. Summary scores are discussed below in the section entitled "Phase I Ratings."

Step 5: Review results. Because the creation of the ratings and the summary scores is almost completely contingent on judgment, a significant amount of decisionmaking is inherent in this approach. A deliberative and iterative process of ranking and review can help to improve confidence in these judgments. The sections below for Phases II through V describe such a process.

The Decision Process

To improve confidence in the summary ranking, an *a priori* decision strategy and decision process need to be agreed upon. Throughout Chapter 4, we have discussed a detailed decision strategy. In this final section of the chapter, we present a decision process that is designed to bring together the previous work establishing standardized measurement and uniform application of those standards. The decision process is composed of five phases.

Phase I. Ratings

In Component 1, health, cost and quality measures were summarized on a common five-point scale centered on 0 (–2, –1, 0, 1, 2), representing strongly negative to strongly positive change. The measures were arrayed in vectors attributable to awardees, groupings, and innovations, shown in Panel A, Table 4.6. In Panel B, those scores may be reassigned to one of three ratings: A = +1, representing significant improvement; B = 0, representing indeterminate improvement; and C = –1, representing significant decline.

Table 4.6. The Ratings System

Panel A—A Ratings System (Component 1)			Panel B—A Decision Strategy (Component 3)
Awardee, Grouping, or Innovation Type	**Measure (health, cost, quality, etc.)**	**Common Scale (–2, –1, 0, +1, +2)**	**A. Significant improvement = +1 B. Indeterminate improvement = 0 C. Significant decline = –1**
Awardee 1	Health measure 1 (e.g., all-cause mortality)	0	A. Significant improvement
	Health measure 2 (e.g., functional status)		
	Cost measure 1 (e.g., program costs)	1	
	Cost measure 2 (e.g., utilization)	-	
	Quality measure 1 (e.g., adverse events)	1	
	Quality measure 2 (e.g., CAHPS)		
Awardee 2	Health measure 1 (e.g., all-cause mortality)	0	C. Significant decline
	Health measure 2 (e.g., functional status)		
	Cost measure 1 (e.g., program costs)	–2	
	Cost measure 2 (e.g., utilization)	-	
	Quality measure 1 (e.g., adverse events)	1	
	Quality measure 2 (e.g., CAHPS)		

Phase II. Initial Prioritization from Cost, Health, and Quality Ratings

The triple aim of the HCIA program focuses awardee attention on lower costs, improved health, and improved quality, which are captured in the second evaluation dimension of program effectiveness. To visualize summary ratings on this dimension, awardees, groups, and interventions of interest could be arrayed on an X-Y-Z scale, where the summary cost rating is X, the summary health rating is Y, and the summary quality rating is Z. Phase I prioritization would involve an initial classification of the awardees, groups, and innovations of interest into three classes: (1) those that can be implemented more broadly, (2) those that need testing in other settings, and (3) those that may be deferred from current consideration for further investment.

Phase III. Validation or Reprioritization with Pooled Analyses and Evidence from Prior Peer-Reviewed Literature

The pooled analyses from Component 2 could then be used to validate scores of priority awardees and interventions or could be used to reprioritize the rankings from Phase II, above. Similarly, CMS could use the rating system to focus its decisionmaking and the attention of the CMS actuary. In some instances, the findings of the ratings system and pooled analyses may differ or be supported by research on the same interventions in the peer-reviewed literature. The relationship of these new findings to prior information from the peer-reviewed literature should be described. The findings of peer-reviewed research may also be considered in prioritization of the rankings from Phase II.

Phase IV. Case Reviews and Evaluation of the Potential for Scale-Up

Summary ratings for each of the quantitative measures and the qualitative measures in the other four dimensions (program implementation, workforce issues, priority populations, and context) could then be compiled into a case review with the goal of identifying why successful interventions were successful and why unsuccessful ones were not. Case reviews can also be used to describe factors that are likely to be important in scale-up of successful interventions. The case reviews and their findings could be reviewed using a panel of reviewers composed of stakeholders or evaluation experts, or both.

Promising innovations need to be assessed for their potential to be scaled up, and projections would be made on the impact of these innovations if carried out nationally. We propose that a simulation model be used to arrive at projections, allowing the evaluator to estimate lower and upper confidence limits for key parameters in the projection and thereby to estimate best- and worst-case estimates following scale-up. These models can be developed in three steps. The first step is to identify the population that may be targeted with the innovation and then to estimate coefficients of cost savings and health gains that result from the innovation. The second step is to identify the per-member, per-month impact of the innovation on spending and health gains. The third is to multiply these benefits by estimates of the population reach that can be expected from nationwide scale-up.

Projection models of cost savings will need to be decomposed into savings accruing to at least two stakeholder groups with a primary interest in cost savings: payers and providers. Decomposition of savings into portions that accrue to these groups will help to determine the extent to which both types of stakeholders may be willing to sustain the interventions over time. Finally, for awardee and grouping innovations that do not result in demonstrated savings, it may be worth using the projection models to explore policy, regulatory, or financial changes that may lead to an improvement in the cost-saving potential of the innovation. A simulation approach to this modeling effort offers the opportunity to estimate potential futures after altering model parameters.

There are several major challenges to the creation of projection models. One such challenge is to define the exact patient population that should be entered into the model. For individual awards, the development of a comparable national patient pool that reflects the award's population is relatively straightforward. However, the inclusion of multiple awards in a synthesis evaluation will complicate projections for a coherent national population because it involves applying benefit estimates from multiple innovation programs conducted on different patient populations. A related challenge is the reality of scale-up: Interventions may not reach 100 percent of their target population nationally if rolled out, and the scaling-up would not necessarily be a linear process.

Conclusions

The Center for Medicare and Medicaid Innovation investment in new care and payment models is of potentially historic importance in terms of how we control health care costs while improving quality and outcomes. The evaluation of these awards will inform decisions about expanding the duration and scope of the models being tested. Despite the challenges, the evaluation and decision process must be of the highest technical quality, as well as transparent and well communicated (Hussey, 2013). CMS has committed to change coupled with an acknowledgment that the solutions that offer promise must be assessed objectively (Shrank, 2013). Thus, evaluators will have a critical role in this effort to reduce costs while maintaining quality in the delivery of health care. The suggestions we have made for evaluations and a decision strategy are put forward with these challenges in mind.

Appendix A. Relationship of Evaluation Dimensions to HCIA Evaluation Questions

Initial planning for implementation of the HCIA programs included substantial consideration of evaluation questions that need to be addressed. These served as the basis for the evaluation approach we outline in this document. In Table A.1, we provide a crosswalk of the HCIA evaluation questions and the relevant evaluation dimensions.

Table A.1. Crosswalk of HCIA Research Questions with Proposed Evaluation Dimensions

Table 3.1 Dimension and Subdimension	Research Questions
II.B	1. Do the models reduce expenditures? Do the models reduce or eliminate variations in expenditures that are not attributable to differences in health status?
II.B.3	a. What changes in expenditures are observed in subpopulations, especially among beneficiaries with the most complex care needs?
I.A. & I.B.1 & I.C	b. How are the models designed to reduce expenditures (e.g., changing the service the population utilizes, reducing the volume or utilization of services, changing the cost of services, etc.)?
II.B.2 & II.D.1	2. Do the models change patterns of utilization? Do the models reduce or eliminate variations in utilization that are not attributable to differences in health status?
II.B.2 & II.D.1	a. If changes in utilization patterns occur, for what types of services and/or populations were these changes observed?
II.D.1 & II.D.2	b. What changes in patterns of utilization are observed in subpopulations, especially among beneficiaries with the most complex care needs?
II.C	3. Do the models provide better quality of care?
II.C	a. If so, how much improvement was seen?
I.B.1 & V	b. Which model characteristics were associated with greater benefit? How can those characteristics be replicated in other settings?
II.B & II.D.2	c. What improvements in quality of care are observed in subpopulations, especially among beneficiaries with the most complex care needs?
II.B & II.D.1	d. To what degree do the model(s) result in reductions in or elimination of disparities in quality of care?
II.A.1	4. Do the models improve health outcomes?
II.A.1 & II.D.2	a. What improvements in outcomes are observed in subpopulations, especially among beneficiaries with the most complex care needs?
II.A.1 & II.D.1	b. To what degree does the model(s) result in reductions in or elimination of disparities in outcomes?

Table 3.1 Dimension and Subdimension	Research Questions
II.D.1	5. Do the models improve or enhance access for beneficiaries?
II.D.1	a. What improvements in access are observed in subpopulations, especially among the most complex beneficiaries?
II.D.1	b. To what degree do the model(s) result in reductions in or elimination of disparities in access?
II.D.1	c. How is this enhanced?
II.D.1	d. Does enhanced access better meet beneficiaries' needs?
II.C.6	6. Do the models provide better coordination of care?
II.C.6 & II.D.2	a. What improvements in coordination of care are observed in subpopulations, especially among the most complex beneficiaries?
I.B.1	b. How is care coordination implemented?
II.B.1	c. What resources are required to provide better coordination of care?
V	d. How can better coordination of care be replicated in other settings?
II.C.3	7. Do the models provide better patient experiences?
II.D.2	a. What improvements are observed in subpopulations, especially among the most complex beneficiaries?
II.D.1	b. To what degree does the model(s) result in reductions in or elimination of disparities in patient experience?
II.D.1	c. What aspects of patient experience are enhanced (e.g., access, perceived care coordination, provider-patient communication, etc.)?
II.D.2	d. Have some models of care inadvertently resulted in worse patient experiences and why?
	8. What factors are associated with the pattern of results (above)? Specifically, are they related to:
I.B.1	a. Characteristics of the models?
V.A	b. Characteristics of the HCIA awardees' approaches to their chosen model (e.g., types and nature of participating providers, utilization of nontraditional types of providers who can interact with patients in their respective communities, specific care coordination interventions used, specific payment incentives, etc.)?
V.A	c. Characteristics of the HCIA awardees' specific features and ability to carry out their proposed intervention?
V.B	d. Characteristics of the HCIA awardees' market or patient populations?
I.B.1	e. Programmatic changes undertaken in response to CMS-sponsored learning and diffusion activities and/or rapid cycle evaluation results?
III	9. How did the models affect the healthcare workforce? To what extent did these effects lead to reductions in expenditures, improvements in quality of care, improvements in outcomes, improvements in access, improvements in care coordination, or improvements in patient experience? Did the models fill health care workforce gaps?
III.A	a. What type of training was provided?
III.A	b. What level of investment in training is required to fill these workforce gaps?
III.A	c. How effective and efficient are the various training models?

Table 3.1 Dimension and Subdimension	Research Questions
I.B.1	10. What changes did models implement to improve quality, access, or care coordination or to lower expenditures?
II.B.1	a. What was the cost in making these changes?
I.C.2	b. How long did it take to implement these changes?
II.B	c. What is the expected cost of sustaining these changes?
1.B.1 & I.B.3	d. What challenges did the participants face in implementing these changes? What lessons were learned from these experiences?
III.C & V.A.4	11. Are the models well received by the practitioners implementing them?
V.A.2	a. How are practitioners involved in the implementation of the model?
V.A.1 & V.A.4	b. How has the organization communicated with practitioners about making changes for the implementation of the model?
V.A.1	c. How involved has the organization's senior management in implementing the model?
V.A.1	d. Has your organization's senior management provided the team that is implementing your model with adequate staff time to carry out tasks related to the implementation?
II.D.3	12. What unintended consequences are observed?
II.D.3	a. What unintended positive consequences were observed?
II.D.3	b. What unintended negative consequences were observed?
II.D.3 & V	c. How could the unintended positive consequences be exploited in future implementation efforts using similar models of care?
II.D.3 & V	d. How could the unintended negative consequences be mitigated in future implementation efforts using similar models of care?

Appendix B. Sample Size Requirements for Estimating HCIA Impacts

We conducted a series of power calculations to generate reasonable estimates of the sample sizes needed to obtain reliable estimates of the of HCIA impacts for on three types of outcomes: (1) medical care utilization, (2) cost of care, and (3) self-reported health status and experience of care. These calculations are intended to inform selections of appropriate evaluation methodologies (i.e., quantitative versus qualitative) and decisions about the potential for efficiency gains through the pooling of data from multiple awards in cases where outcome measures can be standardized.

For each of the three outcomes, we estimated sample size requirements for four types of evaluation designs (see Table B.1 below) that may be considered based on the availability of baseline data and the ability to identify concurrent comparison groups.

Our calculations are intended to reflect sample size requirements for estimating impacts among typical participants in HCIA-sponsored programs—specifically, Medicare and Medicaid beneficiaries in cases where health problems are prevalent and utilization of medical care is frequent. In general, required sample sizes will be higher in the general population where the occurrence of medical care utilization is rarer on average and of more-variable intensity. Further, some awards may have unique designs that require greater sample size (e.g., clustering).

Table B.1. Evaluation Design Descriptions

Evaluation Design	Description
A one-sample comparison relative to a known, fixed benchmark	Comparison of mean utilization measured post-implementation relative to a known, fixed benchmark representing a national standard or value from the published literature
Comparison within the same group pre- and post-intervention (paired comparison)	Pre-post implementation change in mean utilization among a continuously enrolled cohort of program participants (e.g., enrollees in a medical homes demonstration) in the absence of a control group
A two-sample, cross-sectional comparison with a control group (2-sample)	Comparison of mean utilization for distinct groups of treated and untreated program participants (e.g., ICU patients) in the absence of controls
Difference-in-differences comparison	Comparison of pre-post implementation change in mean utilization among program participants to change over the same time period among comparison group of non-program participants

Medical Care Utilization

Our approach assumes that utilization is distributed according to a zero-inflated Poisson distribution in order to account for wide variation in medical care utilization that is due to large numbers of patients with low or no utilization and small number of patients who use large amounts of service. Specifically, we examined the impact of the degree of variation in medical care utilization on sample size required to detect a difference of 5 or 10 percent in baseline utilization by varying the ratio of variance between the empirical distribution of utilization and the Poisson distribution. For example, when the variance inflation is two, the variance of the empirical distribution is twice as large as the standard Poisson or simply two times the sample mean. When the variance inflation is one, the empirical distribution is the same as the Poisson distribution. In Tables B.2 through B.5 below, we show the sample size required to detect a given difference for a range of observed values of mean utilization (e.g., annual number of doctor visits or monthly prescription refills, annual ED visits) for a hypothetical award program that serves Medicare beneficiaries as documented in the 2010 Medical Expenditure Panel Survey.[2] All power calculations shown in the tables below are for two-sided comparison at the significance level of 0.05 and power > 0.80 and selected set of configurations that yield relatively small sample size requirements (roughly < 4,000).

Table B.2. Comparison with a Known, Fixed Benchmark (1-sample comparison)

Benchmark of Mean # Utilizations per Patient	Effect (absolute value)	Variance Inflation	Required Sample Size
2	5%	1	1,572
2	5%	2	3,141
2	10%	1	394
2	10%	2	787
2	10%	4	1,572
5	5%	1	630
5	5%	2	1,258
5	5%	4	2,514
5	10%	1	159
5	10%	2	316
5	10%	4	630
10	5%	1	316
10	5%	2	630
10	5%	4	1,258

[2] Calculated using MEPSnet Query Tools (see Medical Expenditure Panel Survey, undated).

Table B.3. Comparison Within the Same Group Pre and Post Intervention (paired comparison)

Pre-Intervention Mean # Utilizations per Patient	Effect (absolute value)	Variance Inflation	Intra-Class Correlation	Required Sample Size
2	5%	1	0.10	3,460
2	5%	2	0.10	6,910
2	5%	4	0.10	13,820
2	10%	1	0.10	870
2	10%	2	0.10	1,750
2	10%	4	0.10	3,460
5	5%	1	0.10	1,390
5	5%	2	0.10	2,770
5	5%	4	0.10	5,550
5	10%	1	0.10	350
5	10%	2	0.10	700
5	10%	4	0.10	1,390
10	5%	1	0.10	700
10	5%	2	0.10	1,390
10	5%	4	0.10	2,770
10	10%	1	0.10	175
10	10%	2	0.10	350
10	10%	4	0.10	700

Table B.4. Cross-Sectional Comparison with a Control Group (2-sample)

Control Group: Mean # Utilizations per Patient	Effect (absolute value)	Variance Inflation	Required Sample Size
2	5%	1	3,141
2	5%	2	6,280
2	10%	1	786
2	10%	2	1,571
2	10%	4	3,141
5	5%	1	1,257
5	5%	2	2,513
5	5%	4	5,024
5	10%	1	315
5	10%	2	629
5	10%	4	1,257
10	5%	1	629
10	5%	2	1,257
10	5%	4	2,513
10	10%	1	158
10	10%	2	315
10	10%	4	629

Table B.5. Comparison with a Control Group (difference-in-difference)

Control Group: Pre-Intervention Mean Number of Utilizations per Patient	DID Effect (absolute value)*	Variance Inflation	Intra-Class Correlation	Required Sample Size
2	5%	1	0.10	6,950
2	5%	2	0.10	13,850
2	5%	4	0.10	27.700
2	10%	1	0.10	1,750
2	10%	2	0.10	3,500
2	10%	4	0.10	6,950
5	5%	1	0.10	2,800
5	5%	2	0.10	5,550
5	5%	4	0.10	11,100
5	10%	1	0.10	700
5	10%	2	0.10	1,400
5	10%	4	0.10	2,800
10	5%	1	0.10	1,400
10	5%	2	0.10	2,800
10	5%	4	0.10	5,550
10	10%	1	0.10	350
10	10%	2	0.10	700
10	10%	4	0.10	1,400

Medical Care Costs

This series of power calculations is focused on the cost outcomes and based on t-tests of differences in means in a continuous distribution. We used the standard setting of two-sided $p < 0.05$ and power > 0.80, as well as a less conservative setting of one-sided p-value < 0.10 and power > 0.80. We calculated the minimum sample size to detect the corresponding effect size, where the effect sizes were set in based on the empirical distributions in total costs, inpatient costs, and a combined category representing either outpatient or prescription drug costs for a hypothetical population of Medicare beneficiaries, as documented in the 2010 Medical Expenditure Panel Survey.

Table B.6. Comparison with a Fixed Benchmark (1-sample comparison)

Cost Measure	Benchmark Mean/Proportion	SD	Effect Size (% of mean/change in proportion)	Sample Size	Type of Test
Total $ cost	10,000	20,000	1%	> 100,000	2-sided $p < 0.05$
Total $ cost	10,000	20,000	5%	12,560	2-sided $p < 0.05$
Total $ cost	10,000	20,000	10%	3,150	2-sided $p < 0.05$
Total $ cost	10,000	20,000	15%	1,400	2-sided $p < 0.05$
$ inpatient	3,500	13,000	5%	43,350	2-sided $p < 0.05$
$ inpatient	3,500	13,000	10%	10,830	2-sided $p < 0.05$
$ inpatient	3,500	13,000	15%	4,820	2-sided $p < 0.05$
$ inpatient	3,500	13,000	20%	2,710	2-sided $p < 0.05$
$ outpatient or Rx	2,400	7,000	5%	26,800	2-sided $p < 0.05$
$ outpatient or Rx	2,400	7,000	10%	6,680	2-sided $p < 0.05$
$ outpatient or Rx	2,400	7,000	15%	2,970	2-sided $p < 0.05$
$ outpatient or Rx	2,400	7,000	20%	1,680	2-sided $p < 0.05$
Total $ cost	10,000	20,000	1%	> 100,000	1-sided $p < 0.10$
Total $ cost	10,000	20,000	5%	7,250	1-sided $p < 0.10$
Total $ cost	10,000	20,000	10%	1,810	1-sided $p < 0.10$
Total $ cost	10,000	20,000	15%	805	1-sided $p < 0.10$
$ inpatient	3,500	13,000	5%	24,900	1-sided $p < 0.10$
$ inpatient	3,500	13,000	10%	6,220	1-sided $p < 0.10$
$ inpatient	3,500	13,000	15%	2,770	1-sided $p < 0.10$
$ inpatient	3,500	13,000	20%	1,560	1-sided $p < 0.10$
$ outpatient or Rx	2,400	7,000	5%	15,340	1-sided $p < 0.10$
$ outpatient or Rx	2,400	7,000	10%	3,840	1-sided $p < 0.10$
$ outpatient or Rx	2,400	7,000	15%	1,710	1-sided $p < 0.10$
$ outpatient or Rx	2,400	7,000	20%	960	1-sided $p < 0.10$

Table B.7. Comparison Within the Same Group Pre- and Post-Intervention (paired comparison)

Cost Measure	Pretreatment Mean/Proportion	SD	ICC	Effect Size (% change in mean/ proportion)	Required Sample Size	Type of Test
Total $ cost	10,000	20,000	0.1	5%	27,650	2-sided $p < 0.05$
Total $ cost	10,000	20,000	0.1	10%	6,910	2-sided $p < 0.05$
Total $ cost	10,000	20,000	0.1	15%	3,080	2-sided $p < 0.05$
$ inpatient	3,500	13,000	0.1	10%	23,900	2-sided $p < 0.05$
$ inpatient	3,500	13,000	0.1	15%	10,600	2-sided $p < 0.05$
$ inpatient	3,500	13,000	0.1	20%	6,000	2-sided $p < 0.05$
$ outpatient or Rx drug	2,400	7,000	0.1	10%	14,700	2-sided $p < 0.05$
$ outpatient or Rx drug	2,400	7,000	0.1	15%	6,540	2-sided $p < 0.05$
$ outpatient or Rx drug	2,400	7,000	0.1	20%	3,700	2-sided $p < 0.05$
Total $ cost	10,000	20,000	0.1	5%	15,900	1-sided $p < 0.10$
Total $ cost	10,000	20,000	0.1	10%	4,000	1-sided $p < 0.10$
Total $ cost	10,000	20,000	0.1	15%	1,770	1-sided $p < 0.10$
$ inpatient	3,500	13,000	0.1	10%	13,700	1-sided $p < 0.10$
$ inpatient	3,500	13,000	0.1	15%	6,100	1-sided $p < 0.10$
$ inpatient	3,500	13,000	0.1	20%	3,450	1-sided $p < 0.10$
$ outpatient or Rx drug	2,400	7,000	0.1	10%	8,500	1-sided $p < 0.10$
$ outpatient or Rx drug	2,400	7,000	0.1	15%	3,760	1-sided $p < 0.10$
$ outpatient or Rx drug	2,400	7,000	0.1	20%	2,110	1-sided $p < 0.10$

Table B.8. Cross-Sectional Comparison with a Control Group (2-sample)

Cost Measure	Control Mean/Proportion	SD	Effect Size (% change in mean/ proportion)	Required Sample Size	Type of Test
Total $ cost	10,000	20,000	5%	25,200	2-sided $p < 0.05$
Total $ cost	10,000	20,000	10%	6,300	2-sided $p < 0.05$
Total $ cost	10,000	20,000	15%	2,800	2-sided $p < 0.05$
$ inpatient	3,500	13,000	10%	21,700	2-sided $p < 0.05$
$ inpatient	3,500	13,000	15%	9,650	2-sided $p < 0.05$
$ inpatient	3,500	13,000	20%	5,420	2-sided $p < 0.05$
$ outpatient or Rx drug	2,400	7,000	10%	13,500	2-sided $p < 0.05$
$ outpatient or Rx drug	2,400	7,000	15%	5,940	2-sided $p < 0.05$
$ outpatient or Rx drug	2,400	7,000	20%	3,350	2-sided $p < 0.05$
Total $ cost	10,000	20,000	5%	14,450	1-sided $p < 0.10$
Total $ cost	10,000	20,000	10%	3,610	1-sided $p < 0.10$
Total $ cost	10,000	20,000	15%	1,610	1-sided $p < 0.10$
$ inpatient	3,500	13,000	10%	12,450	1-sided $p < 0.10$
$ inpatient	3,500	13,000	15%	5,550	1-sided $p < 0.10$
$ inpatient	3,500	13,000	20%	3,150	1-sided $p < 0.10$
$ outpatient or Rx drug	2,400	7,000	10%	7,700	1-sided $p < 0.10$
$ outpatient or Rx drug	2,400	7,000	15%	3,450	1-sided $p < 0.10$
$ outpatient or Rx drug	2,400	7,000	20%	1,920	1-sided $p < 0.10$

Table B.9. Comparison with a Control Group (difference-in-difference)

Type of Cost	Control Group: Pre-Intervention Mean/Proportion	SD	ICC	DID (absolute value)	Required Sample Size	Type of Test
Total $ cost	10,000	20,000	0.1	5%	55,500	2-sided $p < 0.05$
Total $ cost	10,000	20,000	0.1	10%	13,900	2-sided $p < 0.05$
Total $ cost	10,000	20,000	0.1	15%	6,200	2-sided $p < 0.05$
$ inpatient	3,500	13,000	0.1	10%	47,700	2-sided $p < 0.05$
$ inpatient	3,500	13,000	0.1	15%	21,200	2-sided $p < 0.05$
$ inpatient	3,500	13,000	0.1	20%	12,000	2-sided $p < 0.05$
$ inpatient	3,500	13,000	0.1	25%	7,650	2-sided $p < 0.05$
$ outpatient or Rx drug	2,400	7,000	0.1	10%	29,500	2-sided $p < 0.05$
$ outpatient or Rx drug	2,400	7,000	0.1	15%	13,100	2-sided $p < 0.05$
$ outpatient or Rx drug	2,400	7,000	0.1	20%	7,400	2-sided $p < 0.05$
$ outpatient or Rx drug	2,400	7,000	0.1	25%	4,750	2-sided $p < 0.05$
Total $ cost	10,000	20,000	0.1	5%	31,390	1-sided $p < 0.10$
Total $ cost	10,000	20,000	0.1	10%	7,950	1-sided $p < 0.10$
Total $ cost	10,000	20,000	0.1	15%	3,550	1-sided $p < 0.10$
$ inpatient	3,500	13,000	0.1	10%	27,500	1-sided $p < 0.10$
$ inpatient	3,500	13,000	0.1	15%	12,500	1-sided $p < 0.10$
$ inpatient	3,500	13,000	0.1	20%	6,900	1-sided $p < 0.10$
$ inpatient	3,500	13,000	0.1	25%	4,400	1-sided $p < 0.10$
$ outpatient or Rx drug	2,400	7,000	0.1	10%	16,900	1-sided $p < 0.10$
$ outpatient or Rx drug	2,400	7,000	0.1	15%	7,600	1-sided $p < 0.10$
$ outpatient or Rx drug	2,400	7,000	0.1	20%	4,250	1-sided $p < 0.10$
$ outpatient or Rx drug	2,400	7,000	0.1	25%	2,750	1-sided $p < 0.10$

Other Self-Reported Outcomes

This series of power calculations focuses on continuous patient experience and health outcomes that may be standardized across awardees. We calculated the minimum sample size to detect the corresponding effect size, where the effect sizes were based on published data documenting experiences of adult Medicaid enrollees' experiences with health plans (Chong, Damiano, and Hays, 2012), experiences of adult hospital patients (Hurtado et al., 2005) as assessed using Consumer Assessment of Health Providers and Systems (CAHPS) measures, and self-reported health status using SF-36 scales assessed in the general U.S. population (Ware, undated). All calculations used the standard setting of two-sided $p < 0.05$ and power > 0.80. The power calculation is based on the t-tests for the continuous distribution.

Table B.10. Comparison with a Fixed Benchmark (1-sample comparison)

Type of Self-Reported Outcome	Benchmark Mean	SD	Effect Size (% change in mean)	Required Sample Size
Items in health plan – CAHPS	80	25	5%	310
Items in health plan – CAHPS	80	25	10%	80
Items in hospital – CAHPS	3.5	1	5%	260
Items in hospital – CAHPS	3.5	1	10%	70
Global ratings in hospital – CAHPS	8.5	2	5%	180
Global ratings in hospital – CAHPS	8.5	2	10%	50
SF-36 general health items	75	20	5%	230
SF-36 general health items	75	20	10%	60
SF-36 summary physical or mental component summary	50	10	5%	130
SF-36 summary physical or mental component summary	50	10	10%	35

Table B.11. Comparison Within the Same Group Pre- and Post-Intervention (paired comparison)

Type of Self-Reported Outcome	Pre-Treatment Mean	SD	ICC	Effect Size (% change in mean)	Required Sample Size
Items in health plan – CAHPS	80	25	0.1	5%	680
Items in health plan – CAHPS	80	25	0.1	10%	175
Items in hospital – CAHPS	3.5	1	0.1	5%	570
Items in hospital – CAHPS	3.5	1	0.1	10%	145
Global ratings in hospital – CAHPS	8.5	2	0.1	5%	390
Global ratings in hospital – CAHPS	8.5	2	0.1	10%	100
SF-36 general health items	75	20	0.1	5%	500
SF-36 general health items	75	20	0.1	10%	125
SF-36 summary physical or mental component summary measures	50	0	0.1	5%	280
SF-36 summary physical or mental component summary measures	50	0	0.1	10%	72

Table B.12. Cross-Sectional Comparison with a Control Group (2-sample)

Type of Self-Reported Outcome	Control Group Mean	SD	Effect Size (% change in mean)	Required Sample Size
Items in health plan - CAHPS	80	25	5%	620
Items in health plan – CAHPS	80	25	10%	155
Items in hospital – CAHPS	3.5	1	5%	515
Items in hospital – CAHPS	3.5	1	10%	130
Global ratings in hospital – CAHPS	8.5	2	5%	350
Global ratings in hospital – CAHPS	8.5	2	10%	88
SF-36 general health items	75	20	5%	450
SF-36 general health items	75	20	10%	115
SF-36 summary physical or mental component summary measures	50	10	5%	255
SF-36 summary physical or mental component summary measures	50	10	10%	65

Table B.13. Comparison with a Control Group (difference-in-difference)

Type of Self-Reported Outcome	Pre-Treatment Mean: Control Group	SD	ICC	Effect Size (% change in mean)	Required Sample Size
Items in health plan – CAHPS	80	25	0.1	5%	1,350
Items in health plan – CAHPS	80	25	0.1	10%	340
Items in hospital – CAHPS	3.5	1	0.1	5%	1,130
Items in hospital – CAHPS	3.5	1	0.1	10%	285
Global ratings in hospital – CAHPS	8.5	2	0.1	5%	770
Global ratings in hospital – CAHPS	8.5	2	0.1	10%	200
SF-36 general health items	75	20	0.1	5%	990
SF-36 general health items	75	20	0.1	10%	250
SF-36 summary physical or mental component summary measures	50	10	0.1	5%	555
SF-36 summary physical or mental component summary measures	50	10	0.1	10%	140

References

Atkins D, Eccles M, Flottorp S, et al., "Systems for Grading the Quality of Evidence and the Strength of Recommendations I: Critical Appraisal of Existing Approaches," The GRADE Working Group, *BMC Health Serv Res*, 2004 Dec 22; 4(1):38.

Barber JP, Gallop R, Crits-Christoph P, Frank A, Thase ME, Weiss RD, et al., "The Role of Therapist Adherence, Therapist Competence, and Alliance in Predicting Outcome of Individual Drug Counseling: Results from the National Institute Drug Abuse Collaborative Cocaine Treatment Study," *Psychotherapy Research*, 2006, 16, 229–240.

Blakely C, Meyer J, Gottschalk R, Schmidt N, Davidson W, Roitman D, and Emshoff J, "The Fidelity-Adaptation Debate: Implications for the Implementation of Public Sector Social Programs," *American Journal of Community Psychology*, Vol. 15, No. 3, 1987.

Bullinger M, Andersson R, Cella D, and Aaronsson N, "Developing and Evaluating Cross-Cultural Instruments from Minimum Requirements to Optimal Models," *Quality of Life Research*, 1993; 2:451–9.

Burdine JN, Felix MR, Abel AL, Wiltraut CJ, and Musselman YJ, "The SF-12 as a Population Health Measure: An Exploratory Examination of Potential for Application," *Health Serv Res*, 2000 Oct; 35(4):885–904.

Centers for Medicare & Medicaid Services (CMS), "Evaluation of the Health Care Innovation Awards" (RFA HHSM-500-2012-00093G), 2012.

Centers for Medicare & Medicaid Services (CMS), "Health Care Innovation Awards: Illinois," undated. As of July 29, 2013:
http://innovation.cms.gov/initiatives/Health-Care-Innovation-Awards/Illinois.html

Chong K, Damiano PC, and Hays RD, "Psychometric Performance of the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) 4.0 Adult Health Plan Survey," *Primary Health Care*, 2012, 2:123.

Cleary PD, Lubalin J, Hays RD, Short PF, Edgman-Levitan S, and Sheridan S, "Debating Survey Approaches," *Health Aff*, 1998 Jan; 17(1): 265–66.

Concannon TW, Meissner P, Grunbaum JA, McElwee N, Guise JM, Santa J, Conway PH, Daudelin D, Morrato EH, and Leslie LK, "A New Taxonomy for Stakeholder Engagement in Patient-Centered Outcomes Research," *Journal of General Internal Medicine*, 2012 Jul; 27(8):985–991.

Department of Communities and Local Governments, *Multi-Criteria Analysis: A Manual*, 2009. As of July 31, 2013:
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/7612/1132618.pdf

Deverka PA, Lavallee DC, Desai PJ, Esmail LC, Ramsey SD, Veenstra DL, and Tunis SR, "Stakeholder Participation in Comparative Effectiveness Research: Defining a Framework for Effective Engagement," *J Comp Eff Res*, 2012 Mar; 1(2):181–194.

Fitzpatrick J, "An Introduction to Context and Its Role in Evaluation Practice," *New Directions for Evaluation*, No. 135, Fall 2012.

Funnel S and Roger P, *Purposeful Program Theory: Effective Use of Theories and Change and Logic Models*, Jossey-Bass, 2011.

Glennan T, Bodilly S, Galegher J, and Kerr K, "Introduction: Framing the Problem," in *Expanding the Reach of Education Reforms: Perspectives from Leaders in the Scale-Up of Educational Interventions*, Santa Monica, Calif.: RAND Corporation, MG-248-FF, 2004. As of July 31, 2013:
http://www.rand.org/pubs/monographs/MG248.html

Hartnell CA, Ou AY, and Kinicki A, "Organizational Culture and Organizational Effectiveness: A Meta-Analytic Investigation of the Competing Values Framework's Theoretical Suppositions," *J Appl Psychol*, 2011; 96:677–694.

HCAHPS Online, "HCAHPS Fact Sheet," 2012a. As of January 23, 2012:
http://www.hcahpsonline.org/files/HCAHPS%20Fact%20Sheet%20May%202012.pdf

HCAHPS Online, "Survey Instruments," 2012b. As of January 23, 2012:
http://www.hcahpsonline.org/surveyinstrument.aspx

Helfrich CD, Li YF, Mohr DC, et al., "Assessing an Organizational Culture Instrument Based on the Competing Values Framework: Exploratory and Confirmatory Factor Analyses," *Implement Sci*, 2007; 2:13.

Hurtado MP, Angeles J, Blahut SA, and Hays RD, "Assessment of the Equivalence of the Spanish and English Versions of the CAHPS Hospital Survey on the Quality of Inpatient Care," *Health Serv Res*, 2005 Dec; 40(6 Pt 2):2140–61.

Hussey P, Bankowitz R, Dinneen M, Kelleher D, Matsuoka K, McCannon J, Shrank W, and Saunders R, "From Pilots to Practice: Speeding the Movement of Successful Pilots to Effective Practice," discussion paper, Institute of Medicine, Washington, D.C., 2013. As of July 31, 2013:
http://www.iom.edu/pilotstopractice

Institute of Medicine Committee on Quality of Health Care in America, *To Err Is Human: Building a Safer Health System*, Kohn LT, Corrigan JM, and Donaldson MS, eds., Washington, D.C.: National Academies Press, 2000.

Institute of Medicine Committee on Quality of Health Care in America, *Crossing the Quality Chasm: A New Health System for the 21st Century*, Washington, D.C.: National Academies Press, 2001.

McDonald KM, Sundaram V, Bravata DM, Lewis R, Lin N, Kraft S, McKinnon M, Paguntalan H, and Owens DK, "Care Coordination," Vol. 7 in Shojania KG, McDonald KM, Wachter RM, and Owens DK, eds., *Closing the Quality Gap: A Critical Analysis of Quality*

Improvement Strategies, Technical Review 9, Rockville, Md.: Agency for Healthcare Research and Quality, June 2007.

Medical Expenditure Panel Survey, home page, undated. As of July 30, 2013: http://meps.ahrq.gov/mepsweb/data_stats/meps_query.jsp

Morganti KG, Lovejoy SL, Marsh T, Barcellos SH, Booth-Sutton M, Kase CA, Staplefoote LB, and Berry S, *The CMS Innovation Center Health Care Innovation Awards Database Report: Information on Awardees and their Populations*, unpublished RAND Corporation research, 2013.

Mowbray M, Holter M, Teague G, and Bybee D, "Fidelity Criteria: Development, Measurement, and Validation," *American Journal of Evaluation*, Vol. 24, No. 3, 2003, pp. 315–340.National Research Council (NRC), *Crossing the Quality Chasm: A New Health System for the 21st Century*, Washington, D.C.: The National Academies Press, 2001.

Øvretveit J, Shekelle P, Dy S, McDonald K, Hempel S, Pronovost P, Rubenstein L, Taylor S, Foy R, and Wachter R, "How Does Context Affect Interventions to Improve Patient Safety? An Assessment of Evidence from Studies of Five Patient Safety Practices and Proposals for Research," *BMJ Qual Saf* 2011; 20:604–610.

Owens DK, Lohr KN, Atkins D, et al., "Grading the Strength of a Body of Evidence When Comparing Medical Interventions," *Methods Guide for Comparative Effectiveness Reviews*, Rockville, Md.: Agency for Healthcare Research and Quality, 2009.

Peikes D, Chen A, Schore J, and Brown R, "Effects of Care Coordination on Hospitalization, Quality of Care, and Health Care Expenditures Among Medicare Beneficiaries," *Journal of the American Medical Association*, 2009; 301(6):603.

Sackett D, Rosenberg W, Muir Gray J, Haynes R, and Richardson W, "Evidence-Based Medicine: What It Is and What It Isn't," *British Medical Journal*, 1996; 312:71–2.

Schulte A, Easton J, and Parker J, "Advances in Treatment Integrity Research: Multidisciplinary Perspectives on the Conceptualization, Measurement, and Enhancement of Treatment Integrity," *School Psychology Review*, Vol. 38, No. 4, 2009, pp. 460–475.

Shiely JC, Bayliss MS, Keller SD, et al., *SF-36 Health Survey Annotated Bibliography*, 1st ed., 1998–1995, Boston, Mass.: The Health Institute, New England Medical Centre, 1997.

Shrank W, "The Center for Medicare and Medicaid Innovation's Blueprint for Rapid-Cycle Evaluation of New Care and Payment Models," *Health Affairs*, 32:4, April 2013.

Stecher B and Bohrnstedt G, *Class Size Reduction in California: Summary of the 1998–99 Evaluation Findings*, Santa Monica, Calif.: RAND Corporation, RP-903/1, 2000. As of July 31, 2013: http://www.rand.org/pubs/reprints/RP903z1.html

Stufflebeam DL, "The CIPP Model for Program Evaluation," *Evaluation in Education and Human Services*, Vol. 6, 1983, pp. 117–141.

Turner-Bowker DM, Bayliss MS, Ware JE, Jr., and Kosinski M, "Usefulness of the SF-8 Health Survey for Comparing the Impact of Migraine and Other Conditions," *Qual Life Res* 2003; 12(8): 1003–1012.

Walter S and Brazier JE, "What Is the Relationship Between the Minimally Important Difference and Health State Utility Values? The Case of the SF-6D," *Health and Quality of Life Outcomes*, 2003; 1:4.

Ware JE, "SF-36® Health Survey Update," *The SF Community—SF-36® Health Survey Update*, undated.

Ware JE, Kosinski M, Bayliss MS, et al., "Comparison of Methods for Scoring and Statistical Analysis of SF-36 Health Profile and Summary Measures: Summary of Results from the Medical Outcomes Study," *Medical Care* 1995; 33: AS264–79.

Ware JE, Jr., and Sherbourne CD, "The MOS 36-Item Short-Form Health Survey (SF-36): Conceptual Framework and Item Selection," *Medical Care* 1992; 30(6) 473–83.