

# Evaluating a Community-Based Program to Improve Healthcare Quality: Research Design for the Aligning Forces for Quality Initiative

Dennis P. Scanlon, PhD; Jeffrey A. Alexander, PhD; Jeff Beich, PhD; Jon B. Christianson, PhD; Romana Hasnain-Wynia, PhD; Megan C. McHugh, PhD; Jessica N. Mittler, PhD; Yunfeng Shi, PhD; and Laura J. Bodenschatz, MSW

## Introduction

The Aligning Forces for Quality (AF4Q) initiative, funded by the Robert Wood Johnson Foundation (RWJF), is a multi-site, multi-year initiative with the overarching goals of improving the quality of healthcare and reducing health disparities in its 16 participant communities, and providing models for national reform.<sup>1</sup> Launched in 2006, the AF4Q initiative comprises multiple interventions and goals, which are being developed and revised throughout the program's 10-year course.<sup>2</sup> In addition to being a complex and ambitious initiative in its own right, the AF4Q initiative is being implemented at a time of rapid change in the healthcare arena, when there is a growing awareness of the multiple determinants of health and healthcare quality, and significant national change in healthcare policy.<sup>3,4</sup>

Along with sponsoring the AF4Q initiative, arguably the largest healthcare improvement demonstration project on a community level in the United States to date, the RWJF also dedicates funding to support an independent impartial scientific evaluation of the program. This article provides a description of the research design, data, and limitations of the independent evaluation of the AF4Q initiative. The core purposes of the AF4Q evaluation are to contribute to basic knowledge in 5 main programmatic areas, and answer key questions about its effectiveness in each of these areas: (1) measurement and public reporting; (2) quality improvement; (3) consumer engagement; (4) equity/disparities reduction; and (5) payment reform. Additionally, the evaluation team will answer questions about the effect of aligning the individual programmatic areas. The evaluation team systematically collects evidence related to the facilitators, barriers, and successes of community-based health reform activities and presents it to policy makers, program funders, and communities that are undertaking or contemplating similarly complex initiatives. The AF4Q evaluation also contributes to the health services research literature.

## The AF4Q Initiative: A Complex and Emergent Program

In their seminal work on realistic program evaluation, Pawson and Tilley wrote, "Programs work (have successful 'outcomes')

## Abstract

**Objective:** The Aligning Forces for Quality (AF4Q) initiative is the Robert Wood Johnson Foundation's (RWJF's) signature effort to increase the overall quality of healthcare in targeted communities throughout the country. In addition to sponsoring this 16-site, complex program, the RWJF funds an independent scientific evaluation to support objective research on the initiative's effectiveness and contributions to basic knowledge in 5 core programmatic areas. The research design, data, and challenges faced in the evaluation of this 10-year initiative are discussed.

**Study Design:** A descriptive overview of the evaluation research design for a multi-site, community-based, healthcare quality improvement initiative is provided.

**Methods:** The multiphase research design employed by the evaluation team is discussed.

**Results:** Evaluation provides formative feedback to the RWJF, participants, and other interested audiences in real time; develops approaches to assess innovative and under-studied interventions; furthers the analysis and understanding of effective community-based collaborative work in healthcare; and helps to differentiate the various facilitators, barriers, and contextual dimensions that affect the implementation and outcomes of community-based health interventions.

**Conclusions:** The AF4Q initiative is arguably the largest community-level healthcare improvement demonstration in the United States to date; it is being implemented at a time of rapid change in national healthcare policy. The implementation of large-scale, multi-site initiatives is becoming an increasingly common approach for addressing problems in healthcare. The evaluation research design for the AF4Q initiative, and the lessons learned from its approach, may be valuable to others tasked with evaluating similar community-based initiatives.

*(Am J Manag Care. 2012;18:eS165-eS176)*

For author information and disclosures, see end of text.

only in so far as they introduce the appropriate ideas and opportunities ('mechanisms') to groups in the appropriate social and cultural conditions ('contexts')."<sup>15</sup> As shown in the evaluation team's logic model (Figure), the mechanisms of the AF4Q initiative are a combination of multiple, evolving, targeted interventions; funding; goals for the targeted areas; and various opportunities for technical assistance. Details of this primary logic model, as well as models for the specific program areas, are described in the article by Scanlon et al in this supplement.<sup>2</sup> In general, as the Figure illustrates, the theory of change underlying the AF4Q initiative assumes that alliances (the generic term for the multi-stakeholder partnership in each AF4Q community) will coalesce around designing and implementing programmatic interventions. These interventions are hypothesized to influence intermediate outcomes on the way to improving targeted long-term outcomes. It is expected that it will take some time for the multi-stakeholder initiatives to develop and implement meaningful population-level interventions, and that the timelines will vary for alliances in different communities based on local context and prior experience. As the logic model illustrates, the alliances operate in an environment with many relevant external influences, including changes to state and federal healthcare policy. Complexity in the program is inherent since the RWJF's goal was to establish a dynamic program that would evolve as the work and learning in the participant communities progressed.

The multi-stakeholder alliances charged with implementing the AF4Q interventions operate within a diverse set of communities with distinct social, demographic, and cultural characteristics. These contextual differences suggest that the program interventions may not be homogeneous across participants. The 10-year lifespan of the initiative and other factors related to timing add to the complexity of the evaluation. For example, the initial formation dates of the alliances vary considerably. Not only were communities added during different phases of the AF4Q initiative, some alliances were in existence long before its launch, whereas others were formed or expanded in response to the initiative. Additionally, implementation of the interventions may occur at a different pace in each of the communities for a variety of reasons, including whether any related work was under way in a particular area prior to receiving the AF4Q grant, the level of community engagement around each intervention, and the amount of time each alliance dedicated to achieving community agreement on individual decisions.

### The Evaluation Research Questions

The AF4Q logic model serves as the focal point from

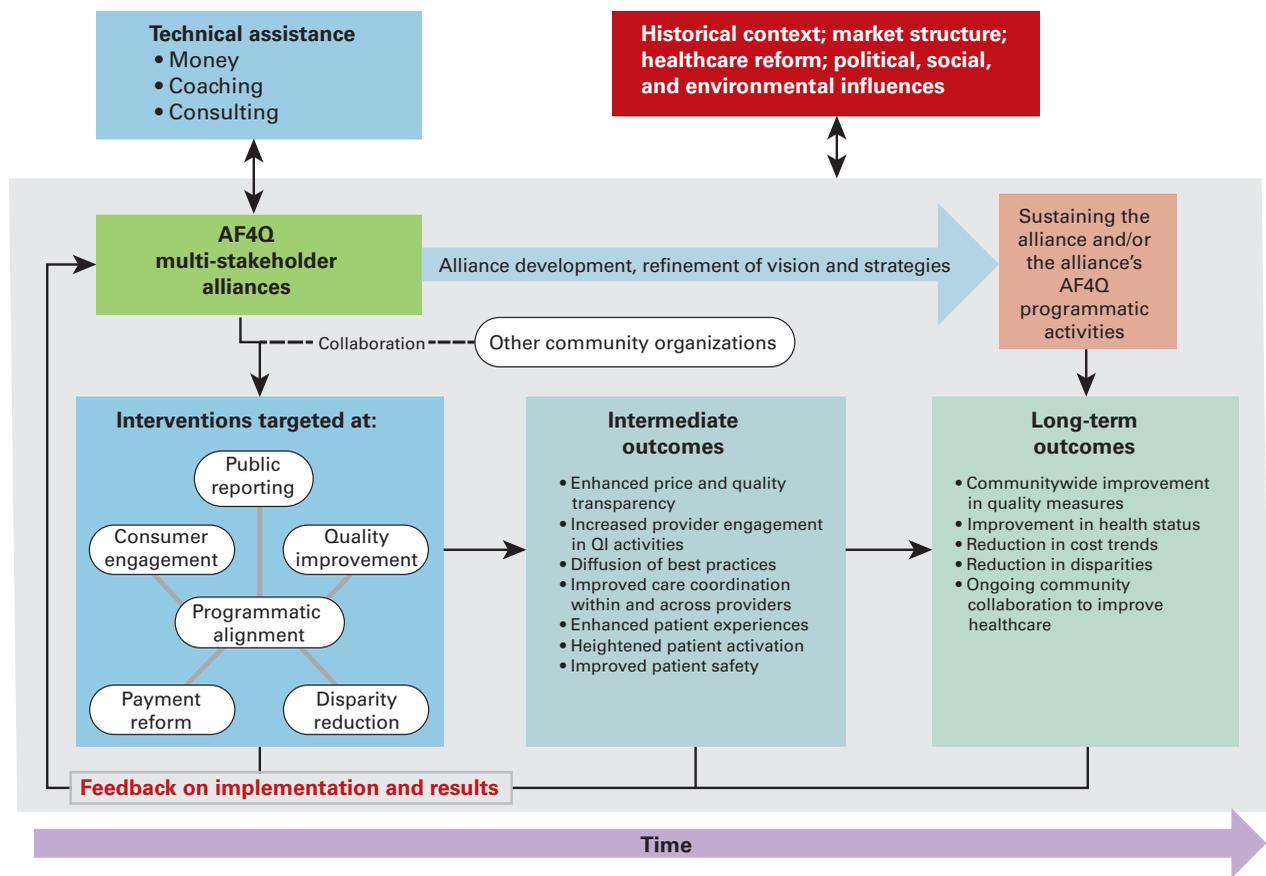
which the research questions, research design, data collection, and data analysis plans for the evaluation are derived. The evaluation team's research goals were purposefully balanced to identify both intermediate and long-term effects of the AF4Q initiative, and to balance the study of progress on these outcomes with an understanding about how and why progress does or does not occur. Thus, the evaluation was designed to include both a summative component, focused on the degree to which expected outcomes are achieved as a result of the program, and a formative component, tasked with developing an ongoing understanding of how the AF4Q initiative unfolds and the expectation that the evaluation team would share information and lessons learned throughout the life of its study. The formative component of the AF4Q evaluation includes frequent and ongoing sharing of information with internal audiences including the RWJF, the AF4Q National Program Office, AF4Q technical assistance providers, and the AF4Q alliances. Formative observations are also shared with external audiences, in the form of research summaries and presentations, and are targeted toward policy makers, healthcare funders, or those in other communities interested in work to improve local healthcare systems.

The AF4Q evaluation research questions are broad and comprehensive. The summative questions focus on the long-term outcomes of the logic model, and the more granular questions focus formatively on the processes used to develop and implement interventions. Some of the research questions also focus on variations in outcomes and processes across alliances, and factors that explain such variations. While the full set of detailed research questions is too voluminous to include in this article, a sample of AF4Q evaluation research questions is included in the Table. The Table organizes research questions according to the relevant component(s) of the logic model and indicates whether a research question is primarily summative or formative in nature.

### The AF4Q Evaluation Research Design

Because of the AF4Q initiative's complex, changing, and voluntary nature, the evaluation team recognized from the outset that a standard experimental design was not appropriate or useful, and that a flexible, rather than fixed, research design was needed. Thus, the interdisciplinary AF4Q evaluation team employs a multiphase research design to study this complex multi-site program. A multiphase design, sometimes referred to as a methodological triangulated design, is one in which there are 2 or more subdesigns. Each subdesign, or "subproject," has independent methodological integrity and

■ **Figure.** Logic Model for the Aligning Forces for Quality Initiative



AF4Q indicates Aligning Forces for Quality; QI, quality improvement.

complements the other(s) to attain the goals of the overall research design.<sup>6,7</sup> Because of the underlying complexity of the program, there are multiple subprojects used in the AF4Q initiative to study the various programmatic areas and their alignment.

As an illustration of how a subproject relates to the overall design, the evaluation team recognized that much of the early work for grantee alliances involved establishing infrastructure, assembling stakeholders, and agreeing on vision and goals. Because of this, the evaluation team focuses not only on the study of the interventions and their effects, but also on the approaches to governance and organization employed in each grantee community. In accordance with its multiphase design, the evaluation team plans to link its findings about programmatic outcomes to the important local governance, organization, and contextual factors identified through its governance-focused work. By bringing these complementary pieces of the overall research design together, the evaluation team is able to assess factors related to the variation in the progress and

effect of the program across AF4Q communities and the programmatic areas.

There are several key research challenges that influenced the choices of the team and were common to all or most portions of the evaluation of the AF4Q initiative. Those challenges are discussed below.

**Changing and Emergent Program**

Despite the high-level clarity that the RWJF had about the types of interventions and people who needed to come together to design and implement them, at the start of the program, detail about how the particular interventions would be designed and implemented was left open for discussion. This approach was important to the RWJF, in that it allowed for in vivo learning and decision processes, and the flexibility to provide more definition in each area when the time was right programmatically. Thus, the evaluation research questions are reviewed periodically and revised or expanded, as needed, to accommodate the evolving program interventions and the rapidly changing local and national healthcare context.

■ **Table.** Sample Research Questions From the Evaluation of the AF4Q Initiative, Organized by Logic Model Component

Logic Model Component	Research Questions (s) summative question (f) formative question
<b>Multi-stakeholder alliances</b>	<ul style="list-style-type: none"> <li>• How do alliances achieve buy-in from multiple stakeholders around AF4Q goals, and how do alliances develop an operational plan for staying focused on and achieving those goals? (f)</li> <li>• How do alliances approach capacity building and sustainability for ongoing long-term collaboration? (f)</li> <li>• What features of alliance organization and governance are related to community quality outcomes or progress toward those outcomes? (s)</li> </ul>
<b>Intervention areas</b> <b>General</b>	<ul style="list-style-type: none"> <li>• What particular activities do alliances select as part of their AF4Q interventions and why do they choose them? (f)</li> <li>• What barriers do alliances encounter when implementing AF4Q interventions? (f) What facilitators help to support the implementation of AF4Q interventions? (f)</li> </ul>
<b>Consumer engagement</b>	<ul style="list-style-type: none"> <li>• What are alliances' goals, strategies, and activities for improving consumer engagement in their communities and how do these efforts meet the AF4Q initiative's expectations? (f)</li> <li>• How do alliances define success and measure progress in consumer engagement, and what are the key influences shaping these expectations for success? (f)</li> <li>• Does consumer engagement improve in AF4Q communities overall? Does improvement in consumer engagement vary by specific behavior or subpopulation? (s)</li> </ul>
<b>Disparities reduction/equity</b>	<ul style="list-style-type: none"> <li>• Are providers (ambulatory and inpatient) in AF4Q communities collecting and using race, ethnicity, and language data to improve quality for populations at risk for receiving low-quality care? (f)</li> <li>• Are alliances successful in collecting and reporting performance data (for physicians and hospitals) disaggregated by racial and ethnic groups? (f)</li> <li>• Are alliances successful in reducing racial and ethnic disparities in quality of care? (s)</li> </ul>
<b>Payment reform</b>	<ul style="list-style-type: none"> <li>• Did AF4Q alliances affect development of new payment approaches in their communities? (s) If so, how? (f)</li> </ul>
<b>Public reporting</b>	<ul style="list-style-type: none"> <li>• Has the AF4Q initiative affected the process of public reporting at the community level? (s) If so, how? (f)</li> <li>• Do stakeholders (consumers, providers, and purchasers) in AF4Q communities respond to public reporting efforts? (s) <ul style="list-style-type: none"> <li>o Have consumers become more aware of public reports and are they more likely to use information in reports for their decision making? (s)</li> <li>o Does the information in public reports influence the quality improvement efforts of community providers? (s)</li> </ul> </li> </ul>
<b>Quality improvement</b>	<ul style="list-style-type: none"> <li>• How do the AF4Q alliances approach the charge of building a sustainable quality improvement infrastructure for the community? How do the alliances build the political and financial support for a community-level approach to quality improvement? (f)</li> <li>• How do external factors (eg, state-based medical home initiatives or the federal government's focus on meaningful use of information technology) affect the alliances' quality improvement efforts? (f)</li> <li>• Do the alliances' quality improvement efforts result in improvements beyond what would be observed if the AF4Q initiative did not exist? (s)</li> </ul>
<b>Programmatic and stakeholder alignment</b>	<ul style="list-style-type: none"> <li>• To what extent are the AF4Q alliances successful in integrating the strategies, activities, and interventions across AF4Q programmatic areas (eg, quality improvement and consumer engagement)? How does this change over time? (f)</li> <li>• To what extent are the AF4Q alliances successful in developing a shared understanding among stakeholders about the alliance's vision and strategies for the AF4Q interventions? How does this change over time? (f)</li> </ul>
<b>Intermediate outcomes</b>	<ul style="list-style-type: none"> <li>• Does transparency about healthcare quality and costs increase in AF4Q communities relative to non-AF4Q communities? (s)</li> </ul>
<b>Long-term outcomes</b>	<ul style="list-style-type: none"> <li>• Does the AF4Q initiative yield improved levels of healthcare quality in participating communities relative to non-AF4Q communities? (s)</li> <li>• Are AF4Q communities able to develop and sustain a communitywide infrastructure to achieve high-quality and affordable healthcare? (s)</li> <li>• Do healthcare disparities decrease in AF4Q communities relative to non-AF4Q communities and do the local healthcare systems promote and support a culture of equity? (s)</li> <li>• Does the level of consumer/patient activation and engagement increase more in AF4Q communities relative to non-AF4Q communities? (s)</li> </ul>

AF4Q indicates Aligning Forces for Quality.

### **Developing Conceptual Frameworks**

One of the main reasons why many of the AF4Q initiative's programmatic areas were not precisely defined by the RWJF at the start of the program was that they were innovative and lacked a strong evidence base. For example, designing and measuring communitywide efforts targeted at engaging consumers to become more active in their health and healthcare is relatively uncharted territory.<sup>8</sup> Because there was no existing model, the evaluation team had to develop a conceptual framework for how consumer engagement might work at the community level to systematically assess the depth and breadth of the work being conducted for consumer engagement. As for other programmatic areas, evaluators were first required to formulate programmatic area-specific logic models and make inferences from empirical data collection about the quality of the processes initiated by communities and the effect of these processes on intermediate and long-term outcomes.<sup>2</sup> More detail on these programmatic area logic models can be found in the article by Scanlon et al in this supplement.<sup>2</sup>

### **Challenges to Measuring Intensity and Scope**

The evaluation team recognized from the outset that because of the size and the multi-faceted and changing nature of the program, it would be difficult to precisely measure the “dose” (or intensity) of each targeted AF4Q intervention, and the initiative at large. The evaluation team also recognized that it would be challenging to systematically measure the relative dose across AF4Q communities, especially since the program was designed by the RWJF to allow each community to have some amount of leeway in how it approached the implementation of each of the interventions.

While the evaluation team employs various approaches to tackling this measurement issue, ranging from general and external (eg, the possibility that AF4Q communities characteristically have been more active in implementing healthcare interventions to date than non-AF4Q communities, and thus could be characterized as such via a binary variable) to more specific and internal to the program (eg, counting and comparing the range of community-level quality improvement activities across AF4Q communities), it is undoubtedly the case that the measurement contains some error, which may create challenges when attempting to link processes to specific program outcomes.

### **Caution Is Needed When Attributing Observed Effects to the AF4Q Initiative**

Because of contextual differences, temporal change, and the complex nature of the program, it is difficult to be defini-

tive when attributing observed outcomes to the AF4Q initiative. To mitigate this concern, the evaluation team uses a variety of data collection and analysis approaches to assess, to the extent possible, the effect of the AF4Q in each programmatic area. When possible, this includes the specification of a control strategy. As discussed in the [Appendix](#), the control group for 2 of our 3 surveys includes a sample of respondents from non-AF4Q areas of the country. For other types of analyses, we attempt to select comparison communities based on population and demographics. Still, any control strategy is imperfect due to the nature of how program participants are selected.

This limitation is common in many program evaluations, and it is important to understand that there is some degree of uncertainty with statements regarding program effects. As reflected in our program logic model (Figure), an additional challenge to attribution is that the types of health improvement work taking place as part of the AF4Q initiative are also taking place, to some degree, in other non-AF4Q communities. For example, there is a national trend toward increased public reporting and transparency of quality measures. While the AF4Q initiative clearly provides resources and a specific structure for this work, these activities are not unique to AF4Q communities. Similarly, implementation of healthcare reform has resulted in many efforts that overlap with the goals of the AF4Q initiative. While evaluators need to be aware of, and account for, temporal trends in their evaluation designs, it is impossible to perfectly control for them.

### **Measuring “Alignment”**

A premise of the initiative is that the absence of synergy, or “alignment” in AF4Q program terms, among key stakeholders and across key programmatic areas (eg, consumer engagement, public reporting) has historically inhibited progress on healthcare quality improvement. From a research perspective, any type of synergy is challenging to define, observe, and measure, because, by definition, it is the interaction of elements that, when combined, produces a total effect that is greater than the contributions of the individual elements.

The evaluation team focuses attention on the measurement of alignment and linking alignment measure(s) to program outcomes. The evaluation team hypothesized, however, that much of the early AF4Q programmatic activity would be focused on individual silos (eg, public reporting, quality improvement, consumer engagement) rather than their alignment, and that this foundational and synergistic component of the AF4Q initiative would not be expected to materialize until later in the program. Because the building

## Reports

of stakeholder alignment is an element of governance and organization, the evaluation team focused early attention on that dimension of synergy. In addition, the evaluation team is employing multiple data collection and measurement strategies to assess the degree to which programmatic alignment is occurring in the overall initiative.

### **Participant Selection and Generalizability**

Another important consideration in study design decisions was that the grantee organizations were not randomly selected. The RWJF chose grantees based upon its own theories about which characteristics of the community and organization are appropriate for desired outcomes.<sup>9</sup>

Consistent with other healthcare programs or community health interventions, the voluntary nature of AF4Q participant community selection can threaten internal and external validity. In many cases, communities were already on their way to implementing key AF4Q-type interventions prior to joining the initiative. Absence of the counterfactual (ie, knowing what would have occurred without the AF4Q initiative) makes it difficult to isolate the true effect of the AF4Q initiative or, at least, the likely effect of randomly selected communities versus those selected competitively.

Acknowledging these threats, the evaluation team strives to clearly communicate the relevant caveats related to both internal and external generalizability when presenting its findings.

### **Formative and Summative Findings**

Because the AF4Q initiative was designed by the RWJF to serve as a demonstration for health improvement programs that addressed complex, real-world issues on a community level, the evaluation team committed from the outset to provide real-time (formative) feedback to the RWJF, its partners, and the participating communities throughout the course of the program. These formative products include presentations about high-level observations, charts and tables that outline grantee approaches or strategies in particular programmatic areas, detailed reports on specific topics or data sets, and results from case studies or from analyses in which 1 or more data sets are summarized and interpreted. To create a reasonable balance between providing formative (ie, throughout the program) and summative (ie, overarching, final) products, the evaluation team continually assesses emergent needs from the RWJF, the communities, and others.

### **The Evaluation Data**

A variety of different sources, including primary and secondary data, are collected and used to answer the research

questions identified for the evaluation in the context of the multiphase design. Importantly, a research approach that combines qualitative and quantitative methods is essential for understanding the effects of the initiative and the processes that comprise the initiative. These sources were designed to be used on a stand-alone basis in some instances, but more often to be purposefully used in combination with other sources to provide contrast and depth to individual analyses, consistent with a methodological triangulated design. The AF4Q evaluation relies on data collected from 3 longitudinal surveys and multiple rounds of interviews with key AF4Q stakeholders, data derived from AF4Q program documentation, and existing observational data collected outside of the AF4Q evaluation. A description of each of the main data sources used in the AF4Q evaluation is available in the Appendix, which also includes details about the purpose and use of each data source, the target population and sampling strategy (where relevant), and other important information.

### **Analytic Approaches Used in Evaluation**

In this section, we describe how data are used to answer the research questions developed for our evaluation of the AF4Q initiative. Our analyses include a variety of single-method quantitative and qualitative approaches; quantitative dominant mixed-method approaches; and qualitative dominant mixed-method approaches. We briefly discuss the approaches in turn, highlighting salient issues.

#### *Quantitative Approaches*

Our quantitative analyses generally rely on survey data collected specifically for the AF4Q initiative (ie, the AF4Q consumer, physician, and alliance surveys) and analysis of secondary data collected by others outside of the AF4Q initiative but relevant for answering specific AF4Q research questions (eg, Dartmouth Atlas of Health Care quality measures). For many of our quantitative analyses, we use a difference-in-difference approach to compare the change in outcomes within the AF4Q community or population of interest (eg, consumer attitudes and opinions) relative to the change for a comparable control group in non-AF4Q communities. The advantage of the difference-in-difference approach is that it washes away any important unobserved confounders that can be considered time invariant. For example, if a specific community has a fixed level of “social capital” that might be important for achieving success on important AF4Q outcomes, then, unlike results in cross-sectional analyses, which might be subject to bias due to important confounders, the results from difference-in-difference

estimates are not threatened due to the longitudinal nature of the research design.

We are not always able to identify a comparable control group, however, and must use different analytic techniques to make inferences. For example, our AF4Q alliance survey asks participants in the multi-stakeholder partnership questions about the leadership, governance, and effect of the AF4Q work. Obviously, we do not observe similar data for non-participants since by definition the relevant questions would not apply to those not participating in an alliance. In this case, we use a longitudinal strategy to examine changes over time among survey respondents, both for the sample of alliance participants at large and the subset of the sample (ie, panel respondents) that provide survey responses at multiple time points. Other analyses undertaken as part of our quantitative research employ cross-sectional analyses to examine associations between key variables and specific outcomes. In addition, we employ descriptive statistics to characterize the distribution of certain variables and highlight the degree of variation within and across the AF4Q communities.

#### *Qualitative Approaches*

The AF4Q logic model and many of the specific research questions developed by the evaluation team focus on understanding how alliances are organized and governed, how the alliances choose to design and implement the various AF4Q programmatic interventions, which factors inform their choices, and what challenges and opportunities alliance stakeholders associate with their participation in the AF4Q initiative. To gain an in-depth understanding of these topics, many of which are focused on processes rather than outcomes of the initiative, the evaluation team collects and analyzes qualitative data in the form of interviews with multiple types of key AF4Q stakeholders and AF4Q program documentation.

Although qualitative data do not lend themselves to generalization and are time-consuming to synthesize and analyze, they are vitally important to the work because they provide comprehensive and detailed information on important processes and meanings that underlie the program. The use of qualitative data also helps the evaluation team paint a more concrete and realistic picture of the evolving AF4Q initiative.

The evaluation team follows a systematic process for all types of key informant interviews that it conducts and identifies multiple respondents to discuss each topic it explores. By gathering data from multiple points of observation, the evaluation team is able to compare and contrast

interviewee viewpoints to gain a dynamic view of issues and processes related to the AF4Q initiative. Interviews are digitally recorded and transcribed in full. The resultant interview transcripts are tagged or “coded” by evaluation team members using deductive high-level (global) categories corresponding to the AF4Q initiative’s main programmatic areas and major concepts that are relevant across all alliances (eg, alliance participation, resources, and structure). These global codes are then entered into a qualitative data analysis software package (Atlas.ti), which allows for large amounts of data in text form to be stored, sorted, and systematically queried. In consultation with the evaluation team’s qualitative data manager, the interview data are pulled from Atlas.ti and analyzed using inductive approaches by evaluation team members working to address specific research questions. Analysis processes conducted by the evaluation team using qualitative data are done using investigator triangulation to minimize bias and ensure systematic analysis of data.

The particular portion of interview data and the analysis processes used in the development of evaluation team products varies depending on several factors: (1) the topic(s) addressed by the research question; (2) the goal of each particular analysis (eg, description, explanation); (3) whether the question can be best addressed by using data from 1 or multiple alliances and/or multiple time points; and (4) the use of other types of data (eg, survey data, documentation) in the analysis and the designated role of each type of data.

#### *Mixed-Methods Approaches*

Some of the questions that the evaluation team seeks to answer can be informed by a combination of the team’s quantitative and qualitative data. While not appropriate for every sub-design within the evaluation, the team has identified questions for which a quantitative or qualitative approach alone is not adequate, such as:

- Did AF4Q alliances affect development of new payment approaches in their communities? If so, how?
- What features of alliance organization and governance are related to community quality outcomes or progress toward those outcomes?

In these situations, the quantitative data are used to assess magnitude and/or frequency of a phenomenon while the qualitative data are used to understand variation in the phenomenon in different contexts, key facilitators and barriers at hand, and/or the meaning that is attributed to

it by those interviewed. In total, the strategic combination of data types in these subprojects allows for a more comprehensive understanding than would be possible with only 1 data type. Additionally, the use of multiple data types and sources allows the evaluation team to compare—or triangulate—findings and look at alternate causes or explanations for findings. While the literature on combining methods and data types is varied, the evaluation team relies on the high-level guidance provided by the National Institutes of Health's recently issued document on mixed-methods research, and as such, understands the importance of making clear choices and designations for the role of all elements that are brought together in analysis processes.<sup>7</sup>

### Conclusion

Because of the myriad of challenges that come with evaluating a multi-site, multi-year initiative intended to solve complex real-world problems, the AF4Q research design is best viewed as a multiphase design of a complex multi-site program. The evaluation seeks to be comprehensive, answer a set of both broad and narrow research questions, and be summative and formative. While the nonrandom, competitive selection of program participants poses threats to internal and external validity, the evaluation team strives to share lessons learned and to disseminate findings, with the appropriate caveats highlighted for consumers of the AF4Q evaluation research.

While there is no exact parallel to the AF4Q initiative, similar large-scale, multi-site initiatives are becoming an increasingly common approach for addressing problems in healthcare. The US Department of Health and Human Services' Chartered Value Exchanges; the Comprehensive Primary Care initiative funded by the Centers for Medicare & Medicaid Services; the Beacon Community Cooperative Agreement Program funded by the Office of the National Coordinator for Health Information Technology; the RWJF's Healthy Kids, Healthy Communities program; and the W.K. Kellogg Foundation's Community-Based Public Health initiative are all examples of contemporaneous programs with strategies similar to those of the AF4Q initiative. These types of programs that are invested in communities as the locus of reform share a common set of characteristics that pose similar opportunities and challenges to researchers studying their effects. We believe that our research design, and the lessons learned from our approach, may be valuable to others tasked with evaluating similar community-based initiatives.

While the findings of the AF4Q evaluation team are not directly generalizable to all communities or populations, the evaluation makes several important contributions. It provides

formative feedback to the program sponsor, participants, and other interested audiences in real time; develops some of the initial research and approaches assessing innovative and under-studied interventions—many of which are now being adopted through national healthcare reform and other large-scale initiatives; furthers the analysis and understanding of effective community-based collaborative work in healthcare; and helps to differentiate the various facilitators, barriers, and contextual dimensions that affect the implementation and outcomes of community-based health interventions.

Specific findings to date from the AF4Q initiative are available through peer-reviewed papers, such as those included in this supplement, and research summaries and special reports. Individuals interested in learning more about the evaluation research design can contact the authors, and those interested in following the findings of the evaluation team can view the full list of AF4Q evaluation products at <http://www.hhdev.psu.edu/chcpr/alignforce>. For more information about the AF4Q initiative and the participating communities, visit [www.forces4quality.org](http://www.forces4quality.org).

**Author affiliations:** Health Management and Policy, School of Public Health, University of Michigan, Ann Arbor, MI (JAA); Jeff Beich Consulting, Grand Island, NY, and Penn State University, University Park, PA (JB); Center for Health Care and Policy Research, Penn State University, University Park, PA (LJB, DPS, YS); Division of Health Policy and Management, University of Minnesota School of Public Health, Minneapolis, MN (JBC); Center for Healthcare Equity and Institute for Healthcare Studies, Division of General Internal Medicine, Northwestern University, Feinberg School of Medicine, Chicago, IL (RH-W); Institute for Healthcare Studies and Department of Emergency Medicine, Northwestern University, Feinberg School of Medicine, Chicago, IL (MCM); Department of Health Policy and Administration, Penn State University, University Park, PA (JNM, DPS).

**Funding source:** This supplement was supported by the Robert Wood Johnson Foundation (RWJF). The Aligning Forces for Quality evaluation is funded by a grant from the RWJF.

**Author disclosures:** Drs Alexander, Beich, Christianson, Hasnain-Wynia, McHugh, Mittler, Scanlon, and Shi and Ms Bodenschatz report no relationship or financial interest with any entity that would pose a conflict of interest with the subject matter of this article.

**Authorship information:** Concept and design (JAA, JB, LJB, JBC, RH-W, MCM, JNM, DPS); acquisition of data (JB, JBC, JNM, DPS); analysis and interpretation of data (JAA, JB, JBC, MCM, JNM, DPS, YS); drafting of the manuscript (LJB, JBC, RH-W, DPS, YS); critical revision of the manuscript for important intellectual content (JAA, JB, LJB, JBC, RH-W, MCM, JNM, DPS, YS); statistical analysis (YS); and obtaining funding (DPS).

**Address correspondence to:** Dennis P. Scanlon, PhD, Department of Health Policy and Administration, Penn State University, 504 Ford Bldg, University Park, PA 16802. E-mail: [dpscanlon@psu.edu](mailto:dpscanlon@psu.edu).

## REFERENCES

1. Quality/equality. Robert Wood Johnson Foundation website. [www.rwjf.org/qualityequality/af4q/index.jsp](http://www.rwjf.org/qualityequality/af4q/index.jsp). Accessed April 17, 2012.
2. Scanlon DP, Beich J, Alexander JA, et al. The Aligning Forces for Quality initiative: background and evolution from 2005 to 2012. *Am J Manag Care*. 2012;18(6):S115-S125.

3. **US Department of Health and Human Services.** Report to Congress: national strategy for quality improvement in health care. <http://www.healthcare.gov/law/resources/reports/quality03212011a.html>. Published March 2011. Accessed July 10, 2012.

4. **Painter MW, Lavizzo-Mourey R.** Aligning Forces for Quality: a program to improve health and health care in communities across the United States. *Health Aff (Millwood)*. 2008;27(5): 1461-1463.

5. **Pawson R, Tilley N.** *Realistic Evaluation*. Los Angeles, CA: Sage Publications; 1997.

6. **Morse J.** Principles of mixed methods and multimethod research design. In: Tashakkori A, Teddlie CB, eds. *Handbook of*

*Mixed Methods Social and Behavioral Research*. Thousand Oaks, CA: Sage Publications; 2003:189-208.

7. **Creswell JW, Klassen AC, Plano Clark VL, Smith KC; for the Office of Behavioral and Social Sciences Research.** Best practices for mixed methods research in the health sciences. [http://obssr.od.nih.gov/mixed\\_methods\\_research/pdf/Best\\_Practices\\_for\\_Mixed\\_Methods\\_Research.pdf](http://obssr.od.nih.gov/mixed_methods_research/pdf/Best_Practices_for_Mixed_Methods_Research.pdf). Accessed July 10, 2012.

8. **Hurley RE, Keenan PS, Martsolf GR, Maeng DD, Scanlon DP.** Early experiences with consumer engagement initiatives to improve chronic care. *Health Aff (Millwood)*. 28(1):277-283.

9. **Aligning Forces for Quality: the Regional Market Project.** 2006 call for proposals. Robert Wood Johnson Foundation website. [http://www.rwjf.org/files/applications/cfp/Aligning\\_Forces\\_CFP.pdf](http://www.rwjf.org/files/applications/cfp/Aligning_Forces_CFP.pdf). Accessed July 17, 2012.

## ■ Appendix. Purpose, Uses, and Descriptions of the Aligning Forces for Quality Evaluation Data

**T**his appendix includes a description of each of the main data sources used in the Aligning Forces for Quality (AF4Q) evaluation, details about the purpose and use of each data source, the target population and sampling strategy (where relevant), and other important information. Additional details on data sources and methods can be obtained by contacting the authors.

### Survey Data

The evaluation team administers 3 surveys to capture important information about the AF4Q initiative, the context in which it operates, and its effects.

#### Consumer Survey

*Purpose and uses:* The consumer survey is designed to capture the components of the AF4Q logic model related to consumer engagement and consumers' use of publicly available quality information. Survey questions focus on patient activation; consumer knowledge of publicly available performance reports that highlight quality differences among physicians, hospitals, and health plans; the ability to be an effective consumer in the context of a physician visit; patient knowledge about their illness; skills and willingness to self-manage the illness; and other related topics.

In order to provide real-time feedback and information to those implementing the AF4Q initiative, and in the spirit of our formative approach, the evaluation team will produce alliance-specific reports of consumer survey results for each of the 3 planned rounds. These reports present the alliance's baseline and longitudinal results and comparisons with other AF4Q communities.

Survey data analysis methods are used to examine distributions of key survey questions, model the variation in responses to survey questions, and identify factors that explain the variation in responses to survey questions. The second round of the

consumer survey data will be used to estimate the effect of the AF4Q initiative on consumer-related outcomes using a difference-in-difference design, where the control group includes a pre- and post-sample of consumers with chronic illnesses drawn from the national comparison sample created from areas of the country that do not include AF4Q communities.

*Target population:* The targeted study population of the consumer survey is adults ( $\geq 18$  years old) with at least 1 of 5 chronic conditions (diabetes, hypertension, heart disease, asthma, and depression). The consumer survey collects data from all of the AF4Q communities and a national comparison sample. The sampling design for the survey is a random digit dialing telephone sample, which was created to yield a representative sample of respondents. Additionally, an over-sample based on respondent race and ethnicity was drawn in 12 of the AF4Q communities to examine differences in survey responses between minorities and non-minorities.

*Additional details:* The consumer survey population was chosen early on in the project, when the AF4Q initiative was focused solely on ambulatory care of individuals with at least 1 of the aforementioned chronic illnesses. Despite the expansion of the AF4Q initiative to include all inpatient care and all members of the population regardless of health status, the consumer survey design has remained the same to provide consistency across the rounds of data collection. Also, because the focus is on those with chronic illness, it ensures that the sample consists of people who are most likely using healthcare services, especially many that are highly relevant to the areas of focus in the AF4Q communities.

#### Physician Survey

*Purpose and uses:* The physician survey (National Survey of Small and Medium-Sized Physician Practices [NSSMPP]) is designed to capture data related to ambulatory quality

## Reports

improvement and assists the evaluation team in learning about the ambulatory QI component of the AF4Q logic model. One of the primary objectives of the NSSMPP is to assess the extent to which physician practices have adopted key components of the chronic care model, the patient-centered medical home, and other care management processes. In addition to organizational information about the practice, the NSSMPP survey instrument includes 7 domains: (1) meaningful use of clinical information technology; (2) use of care management processes to improve the quality of care for 4 chronic diseases (asthma, congestive heart failure, depression, and diabetes); (3) provision of clinical preventive services and health promotion; (4) exposure to external performance incentives such as pay-for-performance and public reporting; (5) payer mix, forms of compensation from health plans, and forms of compensation paid by the practice to its physicians; (6) organizational culture; and (7) information about health plans' provision of care management and preventive services for patients in each practice in the survey.

As with the consumer survey, the evaluation team will produce alliance-specific reports of physician survey results for each of 3 planned rounds.

*Target population:* The NSSMPP collects information about physician practices with 1 to 19 physicians, and because the focus of the NSSMPP is on 4 major chronic diseases, practices were selected only if they were primary care practices; single-specialty cardiology, endocrinology, or pulmonology practices; or multispecialty practices with a significant number of physicians across these specialties. The NSSMPP oversampled the AF4Q communities and insofar as possible, sampled reasonable numbers of practices of each of the above specialty types, and practices in 4 size categories: 1 to 2, 3 to 8, 9 to 12, and 13 to 20 physicians.

*Additional details:* The physician survey was developed by the NSSMPP research team in collaboration with the AF4Q evaluation team, and parallels the National Study of Physician Organizations (NSPO), a longitudinal study of practices with 20 or more physicians, which began in 2001.<sup>1</sup>

The survey was conducted via telephone by a contracted survey firm that interviewed the lead physician or lead administrator of each practice. When this was not possible, the firm interviewed another knowledgeable physician in the practice. Interviews lasted 30 to 45 minutes, and respondents were compensated for their time. The NSSMPP survey was completed in all of the AF4Q communities, and a second

round, renamed the NSPO3, is currently under way; this second round combines the previously separate NSSMPP and NSPO surveys to accommodate a single data collection effort in all practice sizes. A third and final round of the physician survey is planned for 2014 to 2015.

The longitudinal nature of the survey allows for estimates of change over time to identify practice characteristics and market factors that could explain baseline levels and longitudinal changes in practice adoption of quality improvement processes; it also tracks awareness and reaction to public reports of provider quality. A difference-in-difference approach is used to examine the effects in AF4Q communities relative to non-AF4Q communities.

### *Alliance Survey*

*Purpose and uses:* The alliance survey covers the left-hand portion of the AF4Q logic model and connects the relationship of alliance governance and management to programmatic area implementation and program outcomes. It is designed to provide information regarding the degree to which alliance stakeholders are coalescing around a common vision. The survey also allows for assessment of elements of alliance management, leadership, governance, and organizational structure thought to provide the foundation for successful, sustainable collaboration, and demonstrates how these elements change over time.

Similar to the consumer and physician surveys, customized reports are prepared for each AF4Q alliance and provide communities with specific feedback they may use for targeting areas of improvement/attention and identifying success. These reports present baseline and longitudinal results, and comparisons among other AF4Q communities.

*Target population:* The alliance survey targets individuals associated with the alliance as defined by membership on alliance boards, leadership groups, work groups, and staff. Respondents that continue to participate in the alliance are surveyed at multiple times, allowing for comparisons of individual responses over time.

*Additional detail:* The alliance survey is administered online at multiple points throughout the life of the AF4Q initiative. Multiple administrations of the survey, each approximately 18 months apart, facilitate longitudinal comparisons.

Three rounds of the alliance survey were completed in the original AF4Q communities (2006-2012) and at least 1 round was completed in the newer AF4Q communities. By the initiative's end, a total of 5 rounds will be completed

for the original alliances, and based on their program entry dates, 3 or 4 rounds will be completed in the newer AF4Q communities.

### **Qualitative Data**

The evaluation team periodically conducts 3 different types of semi-structured interviews with key stakeholders in the AF4Q alliances: in-person site visit interviews; follow-up telephone interviews; and targeted telephone interviews. Rather than focus on any individual area of the logic model, qualitative data play a role throughout the research design. A high-level description of the 3 types of interviews and the processes used to prepare for interviews is located below; it is followed by a description of the evaluation team's collection and synthesis of documents related to the program.

#### ***In-Person Site Visit Interviews***

To gain the perspective of a variety of stakeholders within each AF4Q community and develop a deep understanding of the alliances' structure and work, the evaluation team periodically conducts site visits. During the 2-day site visits, evaluation team researchers have in-depth, 1-on-1 conversations with a mix of participants in the community. In addition to interviewing alliance staff and volunteer leaders, AF4Q leadership team members, key committee and work group leaders, and other participants in the local AF4Q effort, the evaluation team works to ensure that interviews are conducted with representatives from each of the initiative's targeted community stakeholder groups (eg, consumers, physicians, hospital leaders, healthcare plans, employers, and nurse leaders). The team also identifies 1 or 2 leaders in each community who are not directly involved in the AF4Q initiative to gain an outsider's perspective on the alliance's work. Site visit interview questions are tailored to each interviewee, and a typical interview lasts approximately 1 hour. Collectively, the interviews cover a wide range of topics, including participants' views of the alliance's organizational structure and governance, vision, strategy, collaboration among members, and progress and barriers in each of the AF4Q programmatic areas. The first evaluation team site visit was held approximately 6 months after each community entered the AF4Q initiative, and the second site visit occurred approximately 36 months later in each of the original AF4Q communities. The first site visit was also completed in the newer communities. To date, the evaluation team has conducted a total of 635 in-person site visit interviews resulting in approximately 10,700 pages of double-spaced, typed site visit interview transcripts. Two additional rounds of site visits are planned for each of the AF4Q communities.

#### ***Biannual Phone Interviews With AF4Q Staff Leaders***

The evaluation team also conducts telephone interviews every 6 months with staff leaders in each AF4Q community (eg, AF4Q project directors and/or alliance directors). These 90-minute interviews cover topics such as progress and barriers in each of the AF4Q programmatic areas; changes in alliance governance structure, leadership, and stakeholder participation; the effect of external factors on the alliance's AF4Q efforts; and alliance strategies for alignment of AF4Q programmatic areas. To date, the evaluation team has conducted 8 rounds of interviews with staff leaders resulting in 107 interviews and approximately 2700 pages of double-spaced, typed interview transcripts. The evaluation team plans to continue conducting these interviews for the duration of the AF4Q initiative.

#### ***Targeted Phone Interviews***

Targeted phone interviews complement the site visits and the staff leader interviews by providing an opportunity for in-depth discussions with the individual(s) who lead(s) work in the AF4Q programmatic areas within each alliance and/or AF4Q community. These interviews have been used throughout the study (they were conducted as part of the site visits in the earlier years of the project) and will be conducted annually or semi-annually throughout the remainder of the project. Questions included in these interviews focus on the goals, processes, barriers, and successes in the intervention area of focus.

#### ***Documentation***

The evaluation team gathers, organizes, and synthesizes AF4Q-related documents to understand and track what is happening in each of the AF4Q communities and with the AF4Q initiative on a national level. These data include community funding proposals, information available on alliance or community partner websites, strategic planning documents, meeting agendas and minutes, alliance reports to the AF4Q National Program Office and the Robert Wood Johnson Foundation, news articles and other media, and documents from a host of other sources. In addition, the evaluation team members observe key meetings, webinars, conference calls, and special events to gain additional information. These observations are entered into a projectwide tracking system and provide important context about the program and its implementation.

The documents described here collectively provide the evaluation team with an extensive data set that can be contrasted with the key stakeholder interview data and survey data to challenge conclusions that the team is developing on any given research question. Additionally, documentation

## Reports

in the evaluation team's projectwide tracking system exclusively represents the single-most comprehensive view of the AF4Q initiative from its inception to its current state, and it is used regularly to develop descriptions of the initiative and its evolution.<sup>2</sup>

### Existing Observational Data

Since primary data collection was not possible prior to the start of the AF4Q initiative, the evaluation team uses existing secondary data to explore and understand pre-program trends and ex-ante differences. While these data do not always contain the exact information sought by the evaluation team, they provide valuable insight into these issues. Additional advantages of using these data sources include: (1) cost and collection times are minimized because the data already exist; (2) data that are national in scope and include information on areas outside of the AF4Q alliances; and (3) standardized data which allows for comparable measures across AF4Q alliances and over time. The main secondary data sources used by the evaluation team are described below in greater detail. Other data sources (eg, the US Census, the American Community Survey, and HealthLeaders-InterStudy) that provide descriptive information about the AF4Q communities and their healthcare characteristics are used to supplement our analyses.

#### *Dartmouth Atlas*

The Dartmouth Atlas data contain claims-based quality measures for the fee-for-service Medicare population, computed by AF4Q service areas and other regions not participating in the AF4Q initiative. Specific aspects of quality of care, such as chronic disease management, care coordination, and hospital readmissions, are measured, and serve to assess the AF4Q initiative's effect on long-term quality outcomes identified in the logic model.

#### *Hospital Quality Alliance Program Patient-Level Data*

Over 4200 hospitals voluntarily report their adherence to

recommended processes and treatments for patients admitted for acute myocardial infarction, heart failure, and pneumonia. The Hospital Quality Alliance program data contain hospitals' performance on these process measures, and the associated risk-standardized 30-day readmission and mortality rates. This data source is also used to estimate the effect of the AF4Q initiative on hospital quality over time and reductions in disparities in care relative to hospitals in non-AF4Q communities.

#### *Hospital Consumer Assessment of Healthcare Providers and Systems Survey*

The Hospital Consumer Assessment of Healthcare Providers and Systems Survey provides a standardized instrument and data collection methodology for measuring patients' experience with hospital care. The instrument contains 18 questions that encompass 8 key topics. It is used to assess trends in patient experience with hospital care in AF4Q communities relative to non-AF4Q communities.

#### *The Health Resources and Service Administration's Area Resource File*

The Area Resource File contains measures of resource scarcity and information on health facilities, health professions, health status, economic activity, health training programs, and socioeconomic and environmental characteristics. Data are also available for hospitals in non-AF4Q service areas.

---

## APPENDIX REFERENCES

1. Casalino L, Gillies RR, Shortell SM, et al. External incentives, information technology, and organized processes to improve health care quality for patients with chronic diseases. *JAMA*. 2003;289(4):434-441.
2. Scanlon DP, Beich J, Alexander JA, et al. The Aligning Forces for Quality initiative: background and evolution from 2005 to 2012. *Am J Manag Care*. 2012;18(6):S115-S125.