

# Evaluating a Complex, Multi-Site, Community-Based Program to Improve Healthcare Quality: The Summative Research Design for the Aligning Forces for Quality Initiative

Dennis P. Scanlon, PhD; Laura J. Wolf, MSW; Jeffrey A. Alexander, PhD; Jon B. Christianson, PhD; Jessica Greene, PhD; Muriel Jean-Jacques, MD, MAPP; Megan McHugh, PhD; Yunfeng Shi, PhD; Brigitt Leitzell, MS; and Jocelyn M. Vanderbrink, MHA

**T**he Aligning Forces for Quality (AF4Q) initiative, funded by the Robert Wood Johnson Foundation (RWJF), was a multi-site, multifaceted program with the overarching goals of improving the quality of healthcare and reducing health disparities in its 16 participant communities and providing models for national healthcare reform.<sup>1</sup> Launched in 2006 and concluded in 2015, the AF4Q initiative was built on a community-based multi-stakeholder approach and included multiple interventions and goals, which were developed and revised throughout the program's near decade-long lifespan.

Besides being a complex and ambitious initiative in its own right, the AF4Q program was implemented at a time of rapid change in healthcare marked by a growing awareness of the multiple determinants of health and healthcare quality, and significant national change in healthcare policy, including the passage of the Affordable Care Act (ACA) in 2010.<sup>2,3</sup> A comprehensive overview of the components and phases of the AF4Q program is available in the article by Scanlon et al in this supplement.<sup>4</sup>

In addition to sponsoring the initiative, RWJF dedicated funding to support an independent scientific evaluation of the AF4Q program. The evaluation design included both formative and summative components, which the evaluation team put in place at the beginning of the program and periodically updated in response to the evolution of the program. During the formative phase, the evaluation team focused on developing an ongoing understanding of how the overall program was unfolding by creating subteams to study, in depth, the AF4Q initiative's 5 main programmatic areas (quality improvement, measurement and public reporting, consumer engagement, disparities reduction, and payment reform) and the approaches to governance and organization employed by each grantee community; developing interim findings at multiple points during the program

## Abstract

**Objective:** The Aligning Forces for Quality (AF4Q) initiative was 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, RWJF funded 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 during the summative evaluation phase of this near decade-long program are discussed.

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

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

**Results:** The evaluation team's summative research design involved a data-driven assessment of the effectiveness of the AF4Q program at large, assessments of the impact of AF4Q in the specific programmatic areas, and an assessment of how the AF4Q alliances were positioned for the future at the end of the program.

**Conclusion:** The AF4Q initiative was the largest privately funded community-based healthcare improvement initiative in the United States to date and was 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 summative evaluation research design for the AF4Q initiative, and the lessons learned from its approach, may be valuable to others tasked with evaluating similarly complex community-based initiatives.

*Am J Manag Care. 2016;22:eS8-eS16*

For author information and disclosures, see end of text.

years; and sharing lessons learned during implementation. These findings, along with alliance-specific reports from each of the evaluation team's surveys, provided real-time feedback to internal audiences (ie, RWJF, the AF4Q National Program Office, technical assistance providers, and AF4Q alliances). In addition, formative observations were disseminated to external stakeholders through peer-reviewed publications, research summaries, and presentations. A description of the evaluation design and data sources for the formative phase is located in the article by Scanlon et al.<sup>5</sup>

Approximately 2 years before the AF4Q program ended, the evaluation team began to focus more intently on the summative component, revisiting its initial plan in light of the program's evolution. This paper discusses the team's approach to its summative data collection and analysis and key lessons learned throughout the final phase of this complex multi-site program.

### Summative Evaluation of Complex Programs

The essential purpose of an evaluation is to document what happened in a program, whether changes occurred, and what links exist between a program and the observed impacts.<sup>6,7</sup> As described above in relation to the AF4Q initiative, one way to view evaluations of long-term and complex programs is through 2 interrelated phases: a formative evaluation phase, which produces findings during program implementation, and a summative, or impact evaluation phase, which provides an empirically based appraisal of the final results of a program.<sup>8,9</sup> There is no single format for conducting an evaluation of complex programs like the AF4Q initiative, and the specific approach taken in the design is based on factors such as the characteristics of the program, the requirements of the funding agency, the budget for the evaluation effort, the availability of relevant secondary data sources, and the evaluators' training, research experience, and theoretical lens.<sup>10</sup>

There is, however, a shift occurring in the overall field of evaluation, from focusing on a pre-defined set of program effects for complex initiatives to an emergent approach that aligns more with the multifaceted ways in which social change typically occurs.<sup>11,12</sup> Relatedly, there is an increasing number of evaluations of large-scale and complex community-based programs designed to improve health or healthcare using multimethod and adaptive approaches to study programs and their outcomes.<sup>13-15</sup> Guidance available to researchers on how to approach the evaluation of complex programs in health and health services is also expanding, including a recent article in

the *Journal of the American Medical Association* from CMS, which states that "CMS uses a mixed-methods approach that combines qualitative and quantitative analyses to provide insights into both what the outputs of models are and which contextual factors drive the observed results."<sup>16</sup>

Another key factor to consider in summative evaluation is that funders' interests often go beyond the impact of the particular interventions they sponsored in specific settings to a desire to identify generalizable lessons from the program.<sup>17</sup> RWJF is no exception, describing itself as "passionate about the responsibility it has to share information and foster understanding of the impact of past grant-making—what works, what doesn't, and why."<sup>18</sup> For the AF4Q initiative, RWJF summarized the program goals as, "an unprecedented effort to improve the quality of healthcare in targeted communities, reduce disparities, and provide models to propel reform."<sup>19</sup>

### Developmental Stages of the AF4Q Summative Design

The AF4Q evaluation team's summative design entailed a 3-stage process: putting foundational elements of the summative design in place at the start of the program; closely following the development and changes in the program, and in the larger environment, to account for those factors in the final analysis; and establishing and implementing the final summative design. Each of these stages is described in detail below.

### The Foundations of the Summative Design for the AF4Q Initiative

Overall, the AF4Q evaluation used a multiphase design, formally referred to as a "methodological triangulated design."<sup>20</sup> The design included subprojects with independent methodological integrity in each of the AF4Q initiative's 5 main programmatic areas that were aggregated into an assessment of the whole program during the summative phase.

While the evaluation team agreed that many of the details of the summative work would require adjustments as the program progressed, foundational elements of the summative plan were in place at the beginning. One such element was the plan for a postprogram period to focus on final data collection, analysis, and reporting of the evaluation team's summative findings. Another component involved the early articulation of the overall AF4Q logic model. As Pawson and Tilley wrote in their seminal work on realistic evaluation, "The goal [of evaluation research] has never been to construct theory per se; rather, it has been to develop the theories of practi-

tioners, participants, and policy makers.”<sup>21</sup> Accordingly, the AF4Q logic model, designed to capture the program’s inputs and expected outcomes, was based on RWJF’s plans for the program and the theory of change on which those plans were built. The evaluation team updated the logic model as the program evolved. (A detailed description of the logic model and its development is located in the article by Scanlon et al in this supplement.<sup>4</sup>)

Building from the overall logic model and the more detailed logic models developed for the AF4Q programmatic interventions, the evaluation team developed both formative and summative research questions that corresponded to each programmatic area, the alliance organizations, and the hypothesized intermediate and long-term outcomes of the program. Using those research questions as a guide, the evaluation team then laid out a plan for collecting the needed qualitative and quantitative data (see the online [eAppendix](#) available at [www.ajmc.com](#) for an overview of data). The qualitative data included key informant interviews; observations of principal meetings, discussions, and events; and the compilation of both projectwide and alliance-specific documentation. Quantitative data were obtained from 3 longitudinal surveys (using control groups where possible) and other secondary data sources, such as Medicare claims data from the Dartmouth Group, commercial claims data from MarketScan, and survey data from the Behavioral Risk Factor Surveillance System.

### *Systematic Monitoring of Program and Environmental Change*

During the AF4Q initiative, RWJF and the National Program Office (NPO) enacted many substantial changes to the initiative, including modifications to the requirements in existing programmatic areas, the addition of new programmatic areas, and an expansion of the overall focus of the program, from ambulatory care for people with a specific set of chronic illnesses to all inpatient and outpatient care. (See the article by Scanlon et al in this supplement for a more detailed description of the phases of the AF4Q program.<sup>4</sup>) Monitoring these changes on a program level, and how the alliances adapted to them, became an important aspect of the formative evaluation effort; the team recognized that this work was essential to its ability to assess the program and its outcomes, including the level of alliance fidelity to the interventions and the degree to which observed changes in outcomes could be attributed to the program.

Additionally, although there was only a limited amount of AF4Q-like activity at the start of the program, soon after the AF4Q initiative launched, new programs

that affected many AF4Q alliances were established (eg, the federal government’s Chartered Value Exchange program and the Office of the National Coordinator’s [ONC’s] health information technology projects). Once the ACA was passed in 2010, the pace of change escalated. Dozens of programs aimed at community (and often multi-stakeholder) approaches to improving healthcare quality and cost were launched during the second, third, and fourth funding cycles of the AF4Q initiative. These programs were funded by agencies such as the CMS’s Innovation Center, the Patient-Centered Outcomes Research Institute, the ONC, and others.

The tremendous amount of policy and contextual change meant that attributing observable change to the AF4Q initiative would be even more difficult than anticipated at the outset of the evaluation. To help inform its summative work, the evaluation team followed and documented policy changes at the national level, as well as at the state level for those states with an AF4Q presence. Additionally, the team tracked (ie, followed and documented) alliance involvement in the myriad of health improvement programs that developed during the AF4Q program years. Because of the high volume of environmental change, the evaluation team also conducted “vantage” interviews with a sample of national leaders in health policy to hear perspectives about the AF4Q initiative and its role in the national conversation related to community-based healthcare improvement (see the online [eAppendix](#) available at [www.ajmc.com](#) for details about this effort).

Another major change that the evaluation team monitored was RWJF’s shift to a new strategic direction. During the final 2 years of the AF4Q program, RWJF underwent an internal assessment and emerged with a new priority focused on building a “Culture of Health.”<sup>22</sup> AF4Q leaders and stakeholders in the communities were aware of this shift in RWJF priorities during the final grant period, which was the most flexible in terms of programmatic areas to be addressed in the alliances’ work and in the specific approaches taken by alliances. That flexibility, plus other environmental factors, seemed to have contributed to a number of AF4Q communities’ decisions to explore activities related to population health (in addition to the AF4Q initiative’s traditional areas).

### *Establishing and Implementing the Final Summative Design*

As the formative evaluation work continued in each of the programmatic areas, the evaluation team implemented a process to review its existing plans for the sum-

native phase in light of the program evolution and the environmental changes outlined above. This work began in spring 2013, which was the same time that the alliances were entering into the final 2-year AF4Q funding period. The decisions made by the evaluation team through this process guided the team's final data collection efforts and its approach to the development of summative findings within and across the AF4Q programmatic areas.

Overall, the summative work of the evaluation team involved 5 major overlapping and interconnected processes. The team used regular telephone calls and in-person meetings to provide input and to create linkages across these efforts. Each major summative process is described below.

*Finalizing the Evaluation Team's Summative Goals.* The evaluation team's summative planning process involved a detailed review of the AF4Q logic model, the team's initial summative evaluation plan, the content of the team's papers and reports produced to date, and literature focused on summative evaluation of complex programs. Working from this information, and the experiences and perspectives of individual team members, the team assessed its goals, resources, and available time for the summative effort. As part of this process, the evaluation team sought input from representatives of key audiences: the policy community, the research community, and RWJF itself. Over the course of a 2-day meeting in 2014 with these stakeholders, the evaluation team presented preliminary findings, discussed future plans for the summative work (including envisioned deliverables), and posed questions to representatives from those 3 audiences. Through this feedback process and the evaluation team discussion that followed, the team solidified its goals for the summative evaluation: (1) a data-driven assessment of the effectiveness of the AF4Q program at large, (2) assessments of the impact of the AF4Q initiative in the specific programmatic areas, and (3) an assessment of how the AF4Q alliances were positioned for the future at the end of the program.

*Development of Summative Case Studies for Each AF4Q Alliance.* Recognizing the need to bring individual alliance and overall program level context to its work, the evaluation team began by assembling the contextual information, for a subset of 4 alliances, from a project-wide database collected throughout the evaluation. Following a review and group discussion of that information, the team decided to draft comprehensive, summative case studies for each of the 16 alliances, starting

with the subset of 4 already under discussion. The team divided into 4 case study teams (composed of investigators and research staff) to review existing data and assemble one draft case study per team. The objective of the case studies was to present contextual information about the alliance community, its approach and progress in each of the AF4Q programmatic areas, an overview of governance and organizational changes over time during the AF4Q program, and a summary of the alliance's sustainability planning efforts and its strategic direction.

After the initial 4 draft case studies were assembled, the overall evaluation team held structured discussions during telephone meetings and at an in-person meeting to assess their findings. The goals of the discussions were to identify gaps in knowledge about the alliances, spot points of disagreement among team members about a particular alliance, determine what additional data needed to be incorporated in the case studies, and develop an approach for systematically looking across the program based on the case studies.

From those discussions, the evaluation team developed a set of quantitative "success ratings" that each of the case study teams would use to score their assigned alliance(s). The success ratings, which used a Likert scale, were meant as an aid to organizing and interpreting each team's appraisal drawn from the vast amount of qualitative data compiled about each alliance. They were not intended to be definitive, stand-alone measures of alliance success or progress, or to be externally reported. The draft measure set included items that were specific to the programmatic areas and to broader assessments of the alliance (Table).

The evaluation team had a series of discussions based on the initial 4 case studies and the draft measures of success to identify instances where team members' views about an alliance or a measure of success differed. Because the differences often resulted from team members' varying interpretations of the language, each measure was reviewed, defined, and reworded as needed. Once the evaluation team revised the measures, the case study teams finalized the initial 4 case studies and received assignments to develop case studies for the remaining alliances.

Once the full set of 16 case studies was developed and reviewed by the whole evaluation team, the overall summative evaluation goals were used to guide the design of the final data collection processes. Responsibility for the planning and execution of each final site visit rested with the case study group that drafted the summative case studies for their respective alliances. In doing so, those with the deepest holistic knowledge about an alliance

**Table. Sample Evaluation Team Internal Summative Success Ratings**

Measure	Rating Scale
What is the number of AF4Q program areas the alliance embraced that they had not done work in previously?	0 – The alliance embraced no areas 1 – The alliance embraced 1 new area 2 – The alliance embraced 2 new areas 3 – The alliance embraced 3 new areas 4 – The alliance embraced 4 new areas 5 – The alliance embraced 5 new areas
Due to their involvement in AF4Q, to what degree did the alliance advance the intervention(s) that they were working on prior to AF4Q?	1 – Not at all 2 – A little 3 – Somewhat 4 – A fair amount 5 – To a strong degree NA – Not applicable
The alliance functions as a true multi-stakeholder collaborative, meaning that it has achieved meaningful support, representation, and participation from all sectors (providers, payers, purchasers, and consumers).	1 – Strongly disagree 2 – Disagree 3 – Neither agree nor disagree 4 – Agree 5 – Strongly agree
The alliance plays a critical role in the community, in the eyes of important stakeholders in their community.	1 – Strongly disagree 2 – Disagree 3 – Neither agree nor disagree 4 – Agree 5 – Strongly agree
The interventions implemented by the alliance will sustain and continue within the community after AF4Q.	1 – Strongly disagree 2 – Disagree 3 – Neither agree nor disagree 4 – Agree 5 – Strongly agree
How likely is it that the alliance will sustain itself as a cooperative and continue functioning as an entity after AF4Q?	1 – Very unlikely 2 – Unlikely 3 – Neutral 4 – Likely 5 – Very likely

AF4Q indicates Aligning Forces for Quality.

could tailor protocol questions to fill in gaps in data and determine when additional community-level perspectives about key topics were needed. There were also periodic whole team check-ins throughout the process to maximize the final data collection in ways that supported the evaluation team’s overall goals. Additionally, the case study teams consulted with each programmatic area subteam to ensure that the programmatic area teams could receive the data necessary for their summative efforts. Thus, a customized final site visit plan, which included telephone and in-person interviews and variation in the number of respondents and interview content, was developed specifically for each alliance.

As each final site visit was completed, the case study teams used the new data to finalize the alliance case studies and measures of success scores. All 16 case studies were then

circulated among the whole team for feedback and validation. The iterative production of these 16 case studies was critical to ensuring agreement across all study investigators and research staff before the data were used for analysis.

*Programmatic Summative Findings.* Each programmatic area subteam began to focus on its plans for summative work around the time the final phase of the AF4Q program began in 2013. Starting this process early allowed the subteams the opportunity to identify gaps in data and get feedback on their planned approaches from the evaluation team as a whole. In turn, this allowed the evaluation team to assess how the work of each subteam would fit into the overall summative evaluation.

The programmatic area teams provided data and draft assessments of the alliances’ activities in each area to the case study teams that developed the comprehensive, summative case studies for each alliance. The programmatic area and governance teams also served as a review panel for the success measures related to their respective areas. The programmatic area and governance teams developed specific questions, if necessary, that the case study teams incorporated into the final data collection effort (as described above), and the programmatic area teams were able to use the data from those final site visits to update their work and make their summative assessments of alliance approaches, progress, and outcomes. (Please see

the articles in this supplement on consumer engagement, performance measurement and public reporting, quality improvement, and disparities reduction for details about the findings in each of these programmatic areas; see the sustainability article for an assessment of the end-of-program trajectories of the AF4Q alliances.)<sup>23-27</sup>

*Complete an Assessment of Progress for Each AF4Q Initiative Long-Term Outcome.* One of the components of the AF4Q logic model that was developed early in the program was a list of quantifiable, long-term outcomes. This set of outcomes, selected based on the program’s theory of change and RWJF’s stated goals for the AF4Q initiative, served as the hypothesized set of quantifiable measures for which change in AF4Q communities could be compared with changes in non-AF4Q communities. Early in the program,

the evaluation team identified data sources for each of the long-term outcomes and control strategies to execute a difference-in-difference analysis to determine if the change in these pre-specified outcomes was greater in places that had the AF4Q intervention relative to places that did not. Analyzing some of the long-term outcomes outlined in the logic model depended solely on the team's ability to track changes throughout the program in both the AF4Q and non-AF4Q comparison communities. RWJF eliminated funding for the third (postprogram) round of 2 planned surveys; not having this longitudinal data ultimately precluded the evaluation team from measuring the final program impact on some outcomes.

Another aspect of the summative planning process for long-term outcomes involved the development of a unified framework through which the myriad of outcomes could be presented and discussed. Because of its close fit with the AF4Q programmatic areas and goals, the evaluation team chose to adopt the Triple Aim framework: improving population health, improving quality and experience of care, and reducing cost of care.<sup>28</sup> Details about the analysis approach and findings from the evaluation team's assessment of the AF4Q long-term outcomes can be found in the article by Shi et al in this supplement.<sup>29</sup>

*Conduct a Holistic Assessment of the AF4Q Program.* As the subteams worked on their final assessments of the AF4Q initiative within each of the programmatic areas, the overall evaluation team also began to assess the AF4Q program in its entirety. This analysis took a broad view across the entire AF4Q initiative, reporting aggregate findings based on analyses of all 16 alliances and all major programmatic areas, and incorporated perspectives from the national vantage interviews conducted during the program. Specifically, the evaluation team sought to answer the following research questions through its holistic assessment:

- How should the AF4Q program be viewed in terms of success?
- What important lessons were learned from the AF4Q program that can be useful for those interested in collaborative community-based approaches to improving local healthcare systems and the health of populations residing within these communities?

Exploring answers to these 2 questions involved an iterative, deductive process, with a small group framing an initial discussion and soliciting input and feedback

from the other evaluation team investigators and research staff. Key themes, categorized as dimensions of success or important lessons, were identified individually and then submitted to the whole team for comment and discussion through telephone calls and at an in-person meeting. Based on this feedback, further refinement of these key questions was pursued until consensus was ultimately reached. A full description of the process used, and the findings from this analysis effort, is available in this supplement.<sup>30</sup>

### Challenges and Lessons Learned During the AF4Q Evaluation

Evaluating a multi-site, multifaceted program like the AF4Q initiative comes with a host of challenges. Some issues, such as tracking program and environmental changes, and the need to establish a long-term plan for the design while remaining open to modifications based on emergent factors, were discussed in detail above. Others, including challenges to measuring intensity and scope of the programmatic interventions, limits to attributing observed outcomes to the program, considerations related to data and methods triangulation, and challenges to the generalizability of evaluation findings are discussed in the 2012 paper that focuses on the formative evaluation design for the AF4Q initiative.<sup>5</sup> Limitations that are specific to particular analyses can be found in articles produced as part of the AF4Q evaluation in this supplement and elsewhere.

There are, however, additional challenges to evaluating complex programs that have not been discussed elsewhere in the context of the AF4Q evaluation, such as gaining buy-in and participation in evaluation-related activities from program grantees, navigating the evaluator-funder and other program-level relationships, and coordination and communication issues related to implementing large research projects with team members spread across multiple universities. Because the approaches taken by the AF4Q evaluation team to mitigate some of these common challenges of large-scale evaluations may be of interest to others conducting, funding, or learning from evaluations of complex programs, a brief discussion of each follows.

#### Program Relationships

The importance of the evaluation team's relationships with leaders in each of the participating AF4Q alliances cannot be overstated. While busy implementing the work required of an AF4Q grantee, in addition to conducting work not related to the AF4Q program, the staff of the alliances made time for numerous activities,

## Report

including telephone interviews and site visits, survey participation, responses to requests for documentation, and outreach to their members about the evaluation team's work. The participation of the alliances stems, in part, from the collaborative qualities of the alliance leaders themselves. However, it was also aided by several other factors: RWJF's explicit expectation that grantees' cooperate with the evaluation team (a program requirement), the evaluation team's early outreach efforts to introduce the evaluation design and explain how the team's needs would affect the alliances (these conversations were repeated as alliance leadership changes occurred), the ongoing outreach efforts and relationship building done by team staff and investigators, and the evaluation team's efforts to provide interim findings and data to the alliances throughout the program.

Additionally, ongoing communication to reinforce the role of the evaluation team was important. For example, alliance leaders and members who participated in data collection activities needed to know that the evaluation team had no role in funding decisions and that what the evaluation team learned from individuals would not be reported back to RWJF, the NPO, or others in ways that were identifiable without first gaining explicit permission from the respondent. Another aspect of ongoing communication with the alliance leaders involved asking for their feedback on draft reports and papers. The evaluation team's relationships with RWJF (as the funder of the program and the evaluation) and its agents (ie, the NPO, AF4Q National Advisory Committee, technical assistance providers, and RWJF media consultants), were also important. Strong professional relationships with these groups led to opportunities for the evaluation team to participate in or observe key AF4Q meetings and for information exchange to happen throughout the program years.

Overall, creating and maintaining the space to design and implement an independent scientific evaluation while being responsive to program needs required dedicated time. Much of this work was handled by the AF4Q evaluation principal investigator in conjunction with the RWJF program officer assigned to the evaluation; however, at times, interaction with a broader set of RWJF stakeholders and evaluation team members occurred.

### *Evaluation Team Structure and Process*

The evaluation of a program the size of the AF4Q initiative requires the efforts of a fairly extensive team. Although investigator composition changed over time, 5 or more academic institutions were consistently repre-

sented on the evaluation team. Much of the evaluation team and supporting staff were housed centrally at Penn State, with additional staff members located at the home universities of evaluation investigators.

To implement the day-to-day work of the evaluation, multiple types of subteams were created within the larger evaluation team. As mentioned previously, these subteams included investigators and staff focused on each AF4Q programmatic area and the study of alliance organization and governance. Additionally, there were qualitative and quantitative staff teams, a steering committee composed of the investigators and key staff, and ad hoc teams convened for specific data collection and analysis tasks (eg, case study teams described above). This complexity created the need for well developed processes for communication and feedback across the team that included biweekly steering committee telephone calls and semi-annual in-person meetings of investigators and key staff, multiple weekly subteam and other ad hoc group meetings, daily e-mail and less formal conversations among team members, and a regular, internal evaluation team bulletin that kept team members informed about the myriad of program, alliance, and environmental changes and updates.

Although time consuming and sometimes challenging, the structure and communication processes of the team allowed for investigator triangulation that was critical to the design and adaptability of the evaluation.<sup>31,32</sup> Whereas multiple aspects of the evaluation design benefited from its members' unique set of perspectives, theoretical orientations, and experiences, that same diversity, at times, made for considerable debate and necessitated in-depth and frequent discussions. Thus, another important lesson learned from an evaluation of this size, length, and complexity is that a large portion of the work not only involves data collection, analysis, and writing, but time dedicated to collaboration and communication.

### **Limitations**

Although small within the overall context of health-care, the AF4Q initiative was a large program with a well-funded evaluation compared with most privately funded initiatives. Aspects of the evaluation approach that are outlined here and in the formative evaluation design paper<sup>5</sup> may not be within the scope of evaluations of smaller programs or evaluations of complex programs with more limited funding. Certain aspects of this evaluation design, however, may be adaptable to other programs or contexts, and this description of the multi-dimensionality of the overall evaluation design

may be useful to those in the position of designing the evaluation of larger programs.

Another limitation to adapting the AF4Q evaluation design to other programs is its dedicated focus on a program-level evaluation, rather than each local implementation of the program. In essence, there were 16 different AF4Q programs implemented across the nation. Similar to Weitzman et al in their self-assessment of their evaluation of another RWJF multi-site program, the objective of the programwide evaluation was to not directly assess the impact of the program on any given participant community or identify the particular interventions that were beneficial or detrimental to the aims of individual communities.<sup>15</sup>

Additionally, the external positioning of the independent evaluation team helped to maximize the objectivity of the AF4Q evaluation. This approach is in contrast to the “developmental evaluation” concept that emerged just after the AF4Q initiative launched—an evaluation explicitly designed to be deeply embedded into the program to support the development and evolution of innovations by helping program leaders frame concepts, test intervention iterations, identify unexpected issues, track developments, and monitor changes in context.<sup>33,34</sup> Although the AF4Q evaluation did not follow the developmental approach, the evaluation team did share formative findings and less formal insights with RWJF, the NPO, participant communities, and others throughout the implementation of the program. CMS describes the importance of this type of relationship between program learning and evaluation of complex initiatives as follows: “Placing quantitative assessments in the context of qualitative insights into what is occurring at a given point in a model is critical to program learning and allows model participants to gauge the results of programmatic changes and make adaptations when necessary.”<sup>16</sup>

## Conclusion

Because representatives in both public and private entities at the local, state, and federal levels have dramatically increased their attention to community-based, multi-stakeholder approaches to improving health and healthcare since the AF4Q initiative launched, the design employed by the evaluation team and lessons learned from our approach may be valuable to researchers charged with evaluating similarly complex programs. Additionally, the types of decisions made by the team in the summative phase of the program, and this illustration of design flexibility, may be helpful to evaluation funders as they review proposed research designs. Readers inter-

ested in understanding the full design of the AF4Q evaluation can also read the midterm design article, which provides details about the formative evaluation design and data used during that phase.<sup>5</sup>

Specific findings from the AF4Q initiative are available through peer-reviewed papers, such as those included in this supplement, and previously published research summaries and special reports. Individuals interested in learning more about the AF4Q evaluation 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 [www.hhdev.psu.edu/chcpr/alignforce](http://www.hhdev.psu.edu/chcpr/alignforce). For more information about the participant communities, please see the eAppendix in Scanlon et al<sup>4</sup> in this supplement that includes basic descriptions of the alliances and their website addresses. For archived information about the AF4Q initiative, visit RWJF’s AF4Q website at [www.forces4quality.org](http://www.forces4quality.org).

**Author affiliations:** School of Public Health, The University of Michigan Ann Arbor, MI (JAA); School of Public Health, University of Minnesota Minneapolis, MN (JBC); School of Nursing, George Washington University, Washington, DC (JG); Northwestern University, Feinberg School of Medicine, Division of General Internal Medicine and Geriatrics, Chicago, IL (MJJ); Center for Health Care and Policy Research, Penn State University, University Park, PA (BL, DPS, JMV, LJW); Center for Healthcare Studies, Northwestern University, Feinberg School of Medicine, Chicago, IL (MM); Health Policy and Administration, Penn State University, University Park, PA (DPS, YS).

**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:** Dr Alexander, Dr Christianson, Dr Greene, Dr Jean-Jacques, Ms Leitzell, Dr McHugh, Dr Scanlon, Dr Shi, Ms Vanderbrink, and Ms Wolf report receipt of grants from RWJF. Dr Greene reports meeting or conference attendance on behalf of Insignia Health. Dr Scanlon reports meeting or conference attendance on behalf of RWJF.

**Authorship information:** Concept and design (JAA, JBC, JG, MJJ, MM, DPS, YS, JMV, LJW); acquisition of data (JG, MJJ, DPS, YS, LJW); analysis and interpretation of data (JAA, JG, MJJ, MM, DPS, YS, LJW); drafting of the manuscript (JBC, BL, DPS, LJW); critical revision of the manuscript for important intellectual content (JAA, JBC, JG, MJJ, BL, MM, DPS, YS, JMV, LJW); statistical analysis (YS); obtaining funding (DPS); and administrative, technical, or logistic support (BL, JMV).

**Address correspondence to:** [dpscanlon@psu.edu](mailto:dpscanlon@psu.edu).

## REFERENCES

1. Introduction. Aligning Forces for Quality website. <http://forces4quality.org/node/12010.html>. Accessed July 1, 2016.
2. US Department of Health & Human Services. Report to Congress: national strategy for quality improvement in health care. The Hill website. [thehill.com/images/stories/blogs/quality-strategy.pdf](http://thehill.com/images/stories/blogs/quality-strategy.pdf). Published March 2011. Accessed July 25, 2016.
3. 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. doi: 10.1377/hlthaff.27.5.1461.
4. Scanlon DP, Beich J, Leitzell B, et al. The Aligning Forces for

- Quality initiative: background and evolution from 2005 to 2015. *Am J Manag Care*. 2016;22(suppl 12):S346-S359.
5. Scanlon DP, Alexander JA, Beich J, et al. Evaluating a community-based program to improve healthcare quality: research design for the Aligning Forces for Quality initiative. *Am J Manag Care*. 2012;18(suppl 6):S165-S176.
  6. Balbach ED. Using case studies to do program evaluation. Case Western Reserve University website. [www.case.edu/affil/healthpromotion/ProgramEvaluation.pdf](http://www.case.edu/affil/healthpromotion/ProgramEvaluation.pdf). Published 1999. Accessed July 25, 2016.
  7. Owen JK. *Program Evaluation: Forms and Approaches*. 3rd ed. New York, NY: The Guilford Press; 2007.
  8. Royse D, Thyer BA, Padgett DK, Logan TK. *Program Evaluation: An Introduction*. 3rd ed. Belmont, CA: Wadsworth/Thomson Learning; 2001.
  9. Centers for Disease Control and Prevention. Creating a Culture of Healthy Living: CDC's Healthy Communities Program. CDC website. [www.cdc.gov/nccddphp/dch/programs/healthycommunitiesprogram/tools/pdf/eval\\_planning.pdf](http://www.cdc.gov/nccddphp/dch/programs/healthycommunitiesprogram/tools/pdf/eval_planning.pdf). Accessed July 25, 2016.
  10. Vo AT, Christie CA. Advancing research on evaluation through the study of context. In: Brandon PR, ed. *Research on Evaluation. New Directions for Evaluation*. 2015;148:43-55.
  11. Preskill H, Gopal S, Mack K, Cook J. Evaluating complexity: propositions for improving practice. FSG website. [www.fsg.org/publications/evaluating-complexity](http://www.fsg.org/publications/evaluating-complexity). Published November 2014. Accessed June 27, 2016.
  12. Hargreaves MB. Evaluating system change: a planning guide. Mathematica Policy Research webpage. [www.mathematica-mpr.com/our-publications-and-findings/publications/evaluating-system-change-a-planning-guide](http://www.mathematica-mpr.com/our-publications-and-findings/publications/evaluating-system-change-a-planning-guide). Published April 30, 2010. Accessed July 25, 2016.
  13. Crabtree BF, Nutting PA, Miller WL, Stange KC, Stewart EE, Jaén CR. Summary of the National Demonstration Project and recommendations for the patient-centered medical home. *Ann Fam Med*. 2010;8(suppl 1):S80-S90;S92. doi: 10.1370/afm.1107.
  14. Needleman J, Parkerton PH, Pearson ML, Soban LM, Upenieks VV, Yee T. Overall effect of TCAB on initial participating hospitals. *Am J Nurs*. 2009;109(suppl 11):59-65. doi: 10.1097/01.NAJ.0000362028.00870.e5.
  15. Weitzman BC, Mijanovich T, Silver D, Brecher C. Finding the impact in a messy intervention: using an integrated design to evaluate a comprehensive citywide health initiative. *Am J Eval*. 2009;30(4):495-514. doi: 10.1177/1098214009347555.
  16. Howell BL, Conway PH, Rajkumar R. Guiding principles for Center for Medicare & Medicaid Innovation Model evaluations. *JAMA*. 2015;313(23):2317-2318. doi: 10.1001/jama.2015.2902.
  17. Byrne D. Evaluating complex social interventions in a complex world. *Evaluation*. 2013;19(3):217-228. doi: 10.1177/1356389013495617.
  18. Program Results Reports. Robert Wood Johnson Foundation website. [www.rwjf.org/content/rwjf/en/how-we-work/rel/assessing-our-impact/program-results.html?ref=&k=aligning+forces&st=](http://www.rwjf.org/content/rwjf/en/how-we-work/rel/assessing-our-impact/program-results.html?ref=&k=aligning+forces&st=). Accessed June 28, 2016.
  19. RWJF launches next phase of flagship initiative to lift the quality of American health care. Robert Wood Johnson Foundation website. [www.rwjf.org/qualityequality/product.jsp?id=72287](http://www.rwjf.org/qualityequality/product.jsp?id=72287). Published May 5, 2011. Accessed June 28, 2016.
  20. 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.
  21. Pawson R, Tilley N. *Realistic Evaluation*. Los Angeles, CA: Sage Publications; 1997.
  22. Miller CE, Weiss AF. The view from Aligning Forces to a Culture of Health. *Am J Manag Care*. 2016;22(suppl 12):S333-S336.
  23. Greene J, Farley DC, Christianson JB, Scanlon DP, Shi Y. From rhetoric to reality: consumer engagement in 16 multi-stakeholder alliances. *Am J Manag Care*. 2016;22(suppl 12):S403-S412.
  24. Christianson JB, Shaw BW, Greene J, Scanlon DP. Reporting provider performance: what can be learned from the experience of multi-stakeholder community coalitions? *Am J Manag Care*. 2016;22(suppl 12):S382-S392.
  25. McHugh M, Harvey JB, Hamil J, Scanlon DP. Improving care delivery at the community level: an examination of the AF4Q legacy. *Am J Manag Care*. 2016;22(suppl 12):S393-S402.
  26. Jean-Jacques M, Mahmud Y, Hamil J, Kang R, Duckett P, Yonek JC. Lessons learned about advancing healthcare equity from the Aligning Forces for Quality initiative. *Am J Manag Care*. 2016;22(suppl 12):S413-S422.
  27. Alexander JA, Hearld LR, Wolf LJ, Vanderbrink JM. Aligning Forces for Quality multi-stakeholder healthcare alliances: do they have a sustainable future? *Am J Manag Care*. 2016;22(suppl 12):S423-S436.
  28. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. *Health Aff (Millwood)*. 2008;27(3):759-769. doi: 10.1377/hlthaff.27.3.759.
  29. Shi Y, Scanlon DP, Kang R, et al. The longitudinal impact of Aligning Forces for Quality on measures of population health, quality and experience of care, and cost of care. *Am J Manag Care*. 2016;22(suppl 12):S373-S381.
  30. Scanlon DP, Alexander JA, McHugh M, et al. Summative evaluation results and lessons learned from the Aligning Forces for Quality program. *Am J Manag Care*. 2016;22(suppl 12):S360-S372.
  31. Denzin NK. *Sociological Methods*. New York: McGraw-Hill; 1978.
  32. Patton MQ. Enhancing the quality and credibility of qualitative analysis. *Health Serv Res*. 1999;34(5 pt 2):1189-1208.
  33. Patton MQ. Evaluation for the way we work. *Nonprofit Quarterly*. 2006;13(1):28-33.
  34. Patton MQ. *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York: Guildford Press; 2010.