

The Longitudinal Impact of Aligning Forces for Quality on Measures of Population Health, Quality and Experience of Care, and Cost of Care

Yunfeng Shi, PhD; Dennis P. Scanlon, PhD; Raymond Kang, MA; Megan McHugh, PhD; Jessica Greene, PhD; Jon B. Christianson, PhD; Muriel Jean-Jacques, MD, MAPP; Yasmin Mahmud, MPH; and Jeffrey A. Alexander, PhD

After almost 10 years, the Robert Wood Johnson Foundation's (RWJF's) Aligning Forces for Quality (AF4Q) program, the largest privately funded community-based healthcare initiative to date, ended in April 2015.¹ The AF4Q initiative's design assumed that efforts to improve health and healthcare would be more effective when key stakeholders in a community collaborate to design, coordinate, and implement various interventions aimed at improving health system processes and outcomes.² Between July 2006 and April 2010, 17 multi-stakeholder coalitions (alliances) in communities representing 12.5% of the US population were competitively selected as grantees; 16 were still part of the program when the AF4Q initiative ended. Over the life of the AF4Q program, RWJF dedicated \$300 million in the form of direct grants to participating alliances; payments for technical assistance, program administration, and communication; and funds for an independent evaluation.

The design and implementation of the AF4Q initiative might be characterized by the program's complexity and ambitious scope and the participating communities' diverse geography and distinctive contextual factors, which included small rural areas such as Humboldt County in California; large metropolitan areas such as Boston; and 6 states. As discussed in greater detail in this supplement,¹ the AF4Q interventions, to be implemented by regional multi-stakeholder alliances, were targeted at 5 primary programmatic areas: performance measurement and public reporting, consumer engagement, quality improvement, health equity, and health system payment reform. According to RWJF's goals for the AF4Q program and its theory of change,¹ focused and planned alignment among these different intervention areas would enhance the effectiveness of the initiative in improving population health outcomes, creating more patient-centered care, and improving the value and efficiency in the use of healthcare resources.

Abstract

Objective: To summarize the results from the quantitative analyses conducted during the summative evaluation of the Aligning Forces for Quality (AF4Q) initiative.

Study Design: Longitudinal design using linear difference-in-difference (DD) regression models with fixed effects. Outcomes were selected based on the AF4Q program logic model and organized according to the categories of the Triple Aim: improving population health, improving quality and experience of care, and reducing the cost of care.

Data: Two primary data sources: the AF4Q Consumer Survey and the National Study of Physician Organizations (NSPO); and 4 secondary data sources: the Dartmouth Atlas Medicare claims database, the Truven Health MarketScan commercial claims database, the Behavioral Risk Factor Surveillance System (BRFSS), and the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS).

Results: In total, 144 outcomes were analyzed, 27 were associated with improving population health, 87 were associated with improving care quality and experience, and 30 were associated with reducing the cost of care. Based on the estimated DD coefficients, there is no consistent evidence that AF4Q regions, over the life of the program, showed greater improvement in these measures compared with the rest of the United States. For less than 12% of outcomes (17/144), the AF4Q initiative was associated with a significant positive impact ($P \leq .05$), although the magnitude of the impact was often small. Among the remaining outcomes, with some exceptions, similarly improving trends were observed in both AF4Q and non-AF4Q areas over the period of intervention.

Conclusion and Policy and Practice Implications: Our quantitative findings, which suggest that the AF4Q initiative had less impact than expected, are potentially due to the numerous other efforts to improve healthcare across the United States, including regions outside the AF4Q program over the same period of time. The limited overall impact may also be due to the variability in the "dose" of the interventions across AF4Q regions. However, these results should not be interpreted as a conclusive statement about the AF4Q initiative. More nuanced discussions of the implementation of interventions in the specific AF4Q programmatic areas and their potential success (or lack thereof) in the participating communities are included in other articles in this supplement.

Am J Manag Care. 2016;22:S373-S381
For author information and disclosures, see end of text.

As described by Scanlon et al in an online-only article³ from this supplement, the complexity, scope, and length of the AF4Q initiative created a unique challenge for designing a comprehensive evaluation. Based on the structure and theory of the program, a logic model was developed to provide conceptual links between the various aspects of the program and potentially affected outcomes. The logic model also informed the selection and measurement of outcomes to be studied.¹ The evaluation of the AF4Q initiative employed a mixed-methods approach, and numerous articles were published using quantitative and/or qualitative methods to address specific research questions. Some of these studies focused on a particular programmatic area,^{4,5} while others addressed important issues across different programmatic areas.⁶ In this article, from a summative perspective, we present a unified empirical framework to examine the impact of the AF4Q initiative on a broad set of quantitatively measured outcomes, linked to the 3 important aims in healthcare delivery: improving population health, improving quality and experience of care, and reducing the cost of care.⁷

The Triple Aim was proposed by Berwick et al in 2008,⁷ 2 years after the beginning of the AF4Q initiative; hence, it was not mentioned in the original program design. However, the vision and design of the AF4Q initiative was consistent with the pursuit of the Triple Aim.⁸ More specifically, a central component of the AF4Q program mission was the establishment of local multi-stakeholder alliances operating as the potential “integrators” in selected communities, which has been suggested as a key precondition to achieving progress in the Triple Aim.⁷ As the AF4Q program proceeded, the Triple Aim became increasingly recognized by some key participants as the ultimate goal of such an initiative. Equally important, the outcomes presented in this paper were selected at the outset of the evaluation based on the logic model and the targeted intervention areas in the AF4Q program and before the concept of the Triple Aim initially appeared in the literature. Nevertheless, in retrospect, these preselected quantitative outcomes fit nicely in the Triple Aim framework, which is therefore used to organize our discussion of the results.

Although the quantitative outcomes together may provide an informative and objective synopsis of how much overall progress occurred in the 16 communities regarding the 3 aims, the analysis in this paper was not meant to capture important contextual factors and qualitative components of the AF4Q program. A comprehensive assessment of the AF4Q initiative is presented in

Scanlon et al³ and qualitative assessments of the specific AF4Q programmatic areas can be found in several other papers in this supplement.⁹⁻¹³

The Quantitative Approach in the Summative Evaluation

Overview

The design of the quantitative component of the AF4Q summative evaluation followed 3 principles. First, to avoid data mining, the outcomes and analyses were conceptually driven and determined *ex ante* based on the program logic model. Therefore, these outcomes reflected the commitment of the investigators to the empirical aspect of the evaluation, a practice that has become increasingly important in large-scale evaluation research.¹⁴ Second, all outcomes were tracked longitudinally in the AF4Q regions as well as in other regions included as the comparison sample. Finally, to the extent allowed by the different data sets used in the evaluation, we maintained a consistent difference-in-difference (DD) modeling approach across all outcomes. In total, 144 outcomes were measured and analyzed using 6 different data sources.

Selection of Outcomes

Guided by the AF4Q logic model, we selected quantitative outcomes with several key considerations. First, the whole set of outcomes collectively reflected the potential overall impact of the AF4Q initiative in key areas of health and healthcare delivery. Second, the selected outcomes could be measured quantitatively with a reasonable level of validity and reliability. Third, the outcomes were reasonably well connected to the main programmatic areas, with priorities given to the known focal areas of the alliances’ interventional activities (eg, outcomes related to diabetes and chronic illness). Finally, each AF4Q alliance was required to produce and publicly release reports comparing provider quality in the region. The measures adopted in these public quality reports, often considered important indicators for healthcare quality in the participating communities, were used as references in selecting the outcomes to be evaluated. However, we did not use the actual scores of the quality measures publicly reported by the AF4Q alliances in our analysis because the reported measures, patient populations covered, number of providers included, and the starting time of reporting all varied significantly across the AF4Q alliance communities.

Among the 144 selected outcomes, 27 were associated with improving population health, 87 were associated

■ **Table 1. Characteristics of Data Sources Used in the Study**

Data	Target Population	Unit of Analysis	Data Range	Panel	Information Source
AF4Q Consumer Survey ¹⁵	Chronically ill adults	Individual consumer	2008, 2012	Yes	Survey, primary
National Study of Physician Organizations (NSPO) ¹⁶	Physician practices with ≤20 physicians	Individual physician practice	2009, 2013	Yes	Survey, primary
Dartmouth Atlas ¹⁷	Medicare population	Zip code	2005-2013	Yes	Claims, Medicare
Truven Health MarketScan ¹⁸	Privately insured adult population	Individual health plan enrollee	2006-2013	No	Claims, private insurance
Behavioral Risk Factor Surveillance System (BRFSS) ¹⁹	Adult population	Individual consumer	2007, 2011	No	Survey, secondary
Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) ²⁰	Adult patients admitted to hospitals	Hospital	2008-2013	Yes	Survey, secondary

AF4Q indicates Aligning Forces for Quality.

with improving care quality and experience, and 30 were associated with reducing the cost of care. Each outcome was also categorized based on whether it might exhibit impact in the intermediate or long term, according to the logic model, and was linked to the relevant AF4Q programmatic area(s). The list of all selected outcomes is included in the online [eAppendix](#) (available at www.ajmc.com). To understand the potential change in health equity among different racial and ethnic groups, another programmatic area of the AF4Q initiative,¹² 30 of the 144 outcomes were selected for further analysis (details available upon request).

Data

Two primary data sets collected for this evaluation—the AF4Q Consumer Survey¹⁵ and the National Study of Physician Organizations (NSPO)¹⁶—and 4 secondary data sources—the Dartmouth Atlas Medicare claims database,¹⁷ the Truven Health MarketScan commercial claims database,¹⁸ the Behavioral Risk Factor Surveillance System (BRFSS),¹⁹ and the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS)²⁰—were used to operationalize the 144 outcomes in this analysis. The AF4Q Consumer Survey was a longitudinal survey for chronically ill adults in the 16 AF4Q communities and was completed in 2 waves (August 2008 and November 2012). A comparison sample from the rest of the country was included. The NSPO was also completed in 2 waves (March 2009 and November 2013) and included representative samples of physician organizations from each AF4Q region. A third wave of the AF4Q Consumer Survey and the NSPO was originally planned, but RWJF decided in 2014 not to fund a third wave of either survey. The Dartmouth Atlas data provide zip code-level patient information aggregat-

ed from Medicare claims. The MarketScan data include individual claims from major commercial health plans (Humboldt County, California, was excluded from the AF4Q sample for outcomes using the MarketScan data, which did not allow us to identify single rural counties). The BRFSS is an annual telephone survey conducted by the CDC using repeated cross-sectional samples of individual patients with county identifiers. The HCAHPS annually surveys discharged patients about their recent experience in hospital care. HCAHPS data used in this study were aggregated at the hospital level and did not contain any hospital or patient characteristics. **Table 1** summarizes the main features of the 6 data sets, including the target populations, the units of observation, and the structure and range of the data.

Empirical Strategy

We adopted a unified linear DD approach²¹ to analyzing all of the included outcomes, controlling for fixed effects at the most disaggregate unit allowed by the corresponding data. The baseline regression equation is specified as the following:

$$Y_{it} = \alpha_k + \beta_1 \cdot AFQ + \beta_2 \cdot post + \beta_3 \cdot post \cdot AFQ + X_{it}'\delta + \epsilon_{it}$$

Where Y is the outcome and Xs are the control variables, chosen based on the logic model, information availability in different data sets, and previous empirical studies using similar data. AFQ and post are the indicators for unit i being in an AF4Q region and in a postintervention period, respectively. The (unobserved) fixed effect is captured by α , which is indexed by individual patients in the AF4Q Consumer Survey, zip codes in the Dartmouth Atlas data, counties in the BRFSS, AF4Q regions in the MarketScan data (excluding Humboldt County, California), physician practices in the NSPO,

and hospitals in the HCAHPS. Finally, ϵ is a random error term assumed to have a zero mean conditional on all covariates and the fixed effect. The key parameter of interest is the DD coefficient β_3 , which can be interpreted as the average difference between change in the outcome in the AF4Q regions and the same change in the non-AF4Q regions over the intervention period. Because the AF4Q initiative permitted substantial variation in program implementation by the 16 local multi-stakeholder alliances, the program impact might also differ significantly across communities. To further examine the heterogeneous impact of the AF4Q initiative across regions, we conducted a subsample analysis using the same DD framework to compare each AF4Q region with the non-AF4Q regions.

In all cases, the regression model was estimated by ordinary least square. All binary outcomes were treated as linear, and their DD coefficients could be interpreted as changes in probabilities. Logistic regression was not used in our analysis because the key estimates, the coefficients of interaction terms, do not have easy interpretations under the DD framework.²² The standard errors were adjusted for heteroscedasticity and, when relevant, the serial correlations across repeated observations and the intracluster correlations within the same AF4Q region. Estimated coefficients with $P \leq .05$ are considered statistically significant. Longitudinally applicable population weights were only available in 3 of the 6 data sets. Therefore, to be consistent, all of the results presented here were unweighted. For the 3 datasets (AF4Q Consumer Survey, NSPO, and Dartmouth Atlas data) with usable population weights, we estimated the model with and without weighting, and the results were similar. Finally, for the subset of outcomes selected for analyzing health equity, we extended the above DD approach to account for the differential trends between the white and the minority subpopulations, using a difference-in-difference-in-difference (DDD) model²³ (details available upon request).

Limitations in the Quantitative Approach

There were several important limitations in the analytical approach outlined above. First, the intervention was measured by a binary indicator, which treated the AF4Q initiative as a single dimensional program without variation in “doses” of intervention across different programmatic areas and across different AF4Q alliances. Although this approach might be adequate in providing a simple summary of the overall impact of the program on the whole set of outcomes, the results do not lend

themselves to any interpretation beyond that. However, in the evaluation of a program like the AF4Q initiative, developing refined measurement of intervention “doses” to be included in quantitative analysis poses significant challenges. This is especially true when such measurement requires careful and consistent tracking of a significant amount of information, not only in places exposed to the intervention, but in other areas representing the comparison sample.¹⁵

Second, the comparison sample used in the analysis—all units from non-AF4Q regions included in the corresponding data—was imperfect. Each AF4Q alliance region might be unique in its own fashion, and the combined population of all other regions in the United States may be dissimilar in important characteristics. Methods to construct better comparison groups, such as propensity score matching²⁴ or synthetic control²⁵ were considered. However, the effectiveness of such methods would be limited by the data sets used in our analysis. The AF4Q Consumer Survey and the NSPO have small non-AF4Q samples and only 1 year of pre-intervention data. The Dartmouth Atlas data and the MarketScan data contain only a very small set of basic patient characteristics.

Finally, although our DD model controlled for fixed effects, there are likely other unobserved time-varying factors confounding the results. This is especially a concern because the AF4Q communities were selected competitively rather than randomly, and the program was implemented in an extremely dynamic policy environment, both nationally and locally. However, in the absence of randomization or additional information that can be used to effectively construct instrumental variables, our fixed-effects approach provides conservative and robust estimates to the extent allowed by the data.

Results

Table 2 summarizes the characteristics of the study samples for the pre-intervention and postintervention periods in AF4Q and non-AF4Q regions. All characteristics included in the table also serve as control variables in the estimation of the DD model for the outcomes using that particular data source. In general, the AF4Q and non-AF4Q samples had similar characteristics during the pre-intervention and postintervention periods. A noticeable exception was the NSPO sample, in which the physician practices from the AF4Q regions and those from non-AF4Q regions differ significantly in ownership and type, and this was consistent with what was reported previously.²⁶ Due to

the survey design, the AF4Q Consumer Survey and the NSPO included relatively small non-AF4Q samples used as comparisons.

The online eAppendix (available at www.ajmc.com) reports the unadjusted trends of all the analyzed outcomes in both AF4Q and non-AF4Q regions, as well as the impact of the AF4Q initiative on these outcomes (% points) based on the estimated DD regression coefficients. For most of the outcomes, the AF4Q and non-AF4Q regions showed similar changes over time, and there were no significant differences ($P \leq .05$) between the AF4Q and non-AF4Q regions in the rates of changes. Moreover, when the estimated coefficients suggested significant differences in the rates of changes, the magnitude of the difference was often small. Among the 27 population health outcomes, we found only 1 positive DD coefficient with significance (percentage of adults eating enough fruits or vegetables). However, the positive effect was not due to greater improvement in the AF4Q regions relative to the non-AF4Q regions. Rather, in this case, both AF4Q and non-AF4Q regions showed negative trends. For about 15% of the outcomes in quality and experience of care (13 of 87), the estimated DD coefficients were significantly positive, and most of these positive estimates, except for 1, were due to greater improvements in the AF4Q regions compared with the non-AF4Q regions.

Meanwhile, we found significantly negative DD coefficients for 8 outcomes in this category, due to either less improvement or a larger decrease in these outcomes in the AF4Q regions relative to the non-AF4Q regions. Interestingly, for a number of the outcomes measured separately for the Medicare population and the commercially insured population, our results showed opposite AF4Q effects for these 2 subgroups, although there was no clear pattern.

As for the third aim of reducing the cost of care, our results showed that the total cost of care, and a number of measures of costly care utilization (eg, postdischarge emergency department visits), increased over time in both AF4Q and non-AF4Q regions. These increasing trends are consistent with the findings from other recent studies.²⁷⁻²⁹ Based on the DD

Table 2. Characteristics of Study Sample by Data Sources, Unweighted

Data Source	Pre-intervention		Postintervention	
	AF4Q	Non-AF4Q	AF4Q	Non-AF4Q
AF4Q Consumer Survey				
Average consumers per year (N)	4852	503	4852	503
Medicare, %	43	47	53	56
Medicaid, %	4	5	3	4
Other insurance, %	49	43	39	34
No insurance, %	4	6	5	5
Below poverty level, %	22	28	26	29
Usual source of care: private physician, %	79	77	76	76
Usual source of care: public clinic, %	18	17	18	17
Usual source of care: ED or none, %	4	5	7	7
Diabetes, %	29	30	31	32
Hypertension, %	66	69	59	65
Heart disease, %	16	16	15	18
Asthma, %	17	18	20	22
Depression, %	27	28	32	32
National Study of Physician Organizations (NSPO)				
Average practices per year (N)	626	260	626	260
Practice size (number of physicians)	3.37	4.66	8.57	6.55
Ownership: physicians, %	79	93	73	85
Ownership: medical group, %	4	2	2	1
Ownership: hospital, %	7	3	5	1
Ownership: other, %	10	2	19	14
Practice type: primary care, %	90	41	86	38
Practice type: specialist, %	7	54	7	56
Practice type: multi-specialty, %	4	5	7	6
Revenue from Medicaid, %	11	7	11	8
Revenue from uninsured, %	8	7	9	7
Minority patients, %	26	36	27	40
Patients with limited English proficiency, %	7	8	7	9
Dartmouth Atlas				
Average zip codes per year (N)	3938	35,988	3902	35,629
Female, %	56	55	54	54
Black, %	3	7	4	7

(continued)

Table 2. Characteristics of Study Sample by Data Sources, Unweighted (continued)

Data Source	Pre-intervention		Postintervention	
	AF4Q	Non-AF4Q	AF4Q	Non-AF4Q
Dual-eligible, %	10	15	11	14
Age: 65-69 years, %	22	23	24	25
Age: 70-74 years, %	24	25	24	25
Age: 75-79 years, %	21	21	19	19
Age: 80-84 years, %	16	16	15	14
Age: 85-99 years, %	16	15	18	16
Truven Health MarketScan				
Average enrollees per year (N)	88,615	806,982	216,726	1,999,226
Female, %	49	49	47	48
Age: 18-25 years, %	2	1	2	2
Age: 26-35 years, %	5	5	5	6
Age: 36-50 years, %	29	32	29	32
Age: 51-64 years, %	65	62	64	61
Comorbidity (Charlson Comorbidity Index)	1.16	1.09	1.13	1.12
Behavioral Risk Factor Surveillance System (BRFSS)				
Average consumers per year (N)	51,319	358,320	67,567	455,417
Female, %	62	63	59	61
Age, mean	53.59	54.11	54.67	55.41
White, %	86	76	85	75
Black, %	7	9	6	10
Hispanic, %	4	8	4	8
Asian, %	2	2	2	2
Other race, %	2	5	3	4
Education: elementary school or less, %	2	4	2	3
Education: some high school, %	5	7	4	6
Education: high school degree, %	27	31	27	30
Education: some college, %	27	26	27	27
Education: college degree or more, %	39	32	40	33
Health insurance: yes (%)	91	88	91	88
Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) ^a				
Average hospitals per year (N)	339	2251	453	3336

AF4Q indicates Aligning Forces for Quality; ED, emergency department.
^aHospital or patient characteristics were not available in the version of HCAHPS used here.

coefficients, the total cost of care per enrollee in the Medicare population increased less in the AF4Q regions than in the non-AF4Q regions, by 1 percentage point. However, in the commercially insured population, the same measure increased more in the AF4Q regions than in the non-AF4Q regions, by 2 percentage points. For 10 measures of costly care utilization, the AF4Q regions increased significantly more (or decreased less) than the non-AF4Q regions, although the magnitude of the difference, with only 1 exception, was 2 percentage points or less.

The Figure shows, for each aim and individual AF4Q region, the number of outcomes with greater, similar, or less improvement relative to the non-AF4Q sample, based on the significance ($P \leq .05$) of the estimated DD coefficients of all 144 outcomes. (Humboldt County, California had a smaller number of outcomes, as it was excluded from the analysis using the MarketScan data.) A similar approach to visualizing DD estimates for multiple outcomes was used in a previous study.³⁰

The estimated AF4Q effects on individual participating regions showed that the patterns in the majority of the AF4Q regions were similar to the overall pattern, although there was some variation across these regions in terms of the number of significantly impacted outcomes. There seemed to be more cross-region variation in the number of cost measures affected by the AF4Q initiative, although the size of the effect was generally small. Finally, based on the results from our analysis of the health equity outcomes (available upon request), we found no evidence that the AF4Q initiative was effective in reducing disparities among different racial and ethnic groups in terms of the selected outcomes.

Discussion

With a few exceptions, we did not find consistent evidence that AF4Q regions, over the life of the program, improved population health, improved quality and experience of care, or reduced the cost of care more than did the rest of the United States. Our results suggest that while the majority of outcomes measuring quality and experience of care in the

AF4Q regions improved, these outcomes also improved in other regions during the same period and often with similar magnitudes. This might be due to other significant efforts to improve care quality in regions outside AF4Q over the same period of time.

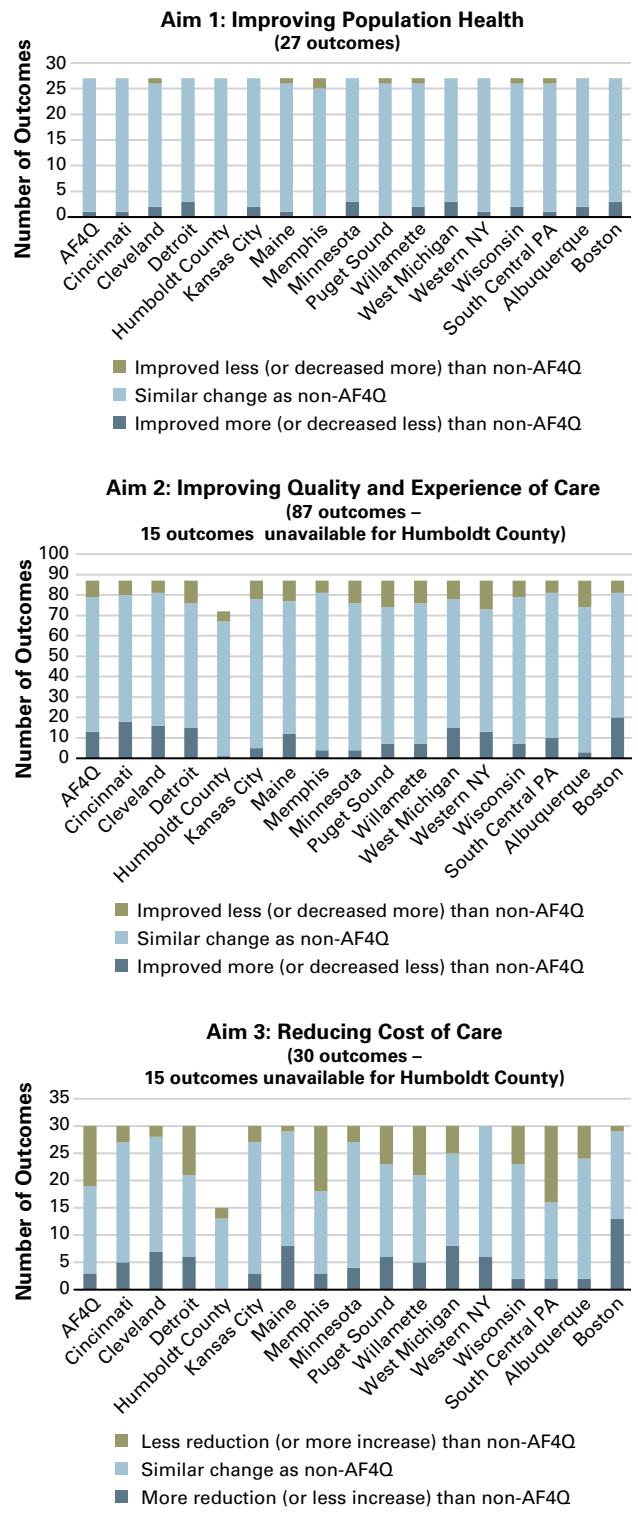
In addition, some AF4Q interventions might not have been effectively implemented or might have been implemented at a scale not large enough to impact the targeted population. Especially in the programmatic areas of consumer engagement and health equity, the success in implementing interventions was limited.^{11,12} On the other hand, a large number of the outcomes measuring population health and cost of care did not improve or became worse in both AF4Q and non-AF4Q regions over the study period. This may suggest the particular challenges in improving population health and reducing cost of care.

We should also note that 5 of the 6 data sets used in this analysis had reasonably large samples and enough power to detect meaningful impact at the population level. (The only exception might be the NSPO.) Based on the confidence intervals, for most of the estimated coefficients without significance ($P > .05$), empirically meaningful effects can be ruled out.

It is worth pointing out that the small number of outcomes that did show a positive AF4Q impact were typically intermediate outcomes measuring processes of chronic illness care delivery (eg, cholesterol screening rates). Such outcomes were often more directly targeted in the alliances' interventions and might take relatively less time to improve. However, even when a significant effect was found, the magnitude was often small. This seemed to be consistent with the observed trends in the similar measures publicly reported in these regions.

We also found that the impact of the AF4Q initiative varied across the participating communities, especially in outcomes related to cost of care. Such heterogeneity might be explained by different implementation of the main programmatic areas of intervention and the important contextual factors in each region that are not controlled in this analysis. The Figure, although an easily understood graphical synopsis across all AF4Q regions, might not fully capture the cross-region heterogeneity in the program impact. This is because the Figure does not distinguish between different outcomes within the same category or account for the magnitude of impact. For example, compared with non-AF4Q regions, 2 AF4Q regions might both show greater improvement in 5 outcomes measuring quality and experience of care, but

Figure. Number of Outcomes Showing Significantly ($P \leq .05$) Greater, Less, or Similar Improvement, Relative to the Non-AF4Q Regions, Organized by Aim and AF4Q Community



AF4Q indicates Aligning Forces for Quality.

these could be 5 different outcomes with different magnitudes of improvement.

The difference in the patterns of AF4Q impacts between the commercially insured population and the Medicare population, as shown in a number of outcomes, might be due to particular interventions and policy environments that differentially affected the 2 subgroups in these regions. More nuanced discussions of the implementation of specific interventions in the AF4Q programmatic areas and their potential success (or lack thereof) in the participant communities can be found in other articles in this supplement.⁹⁻¹²

Finally, although the quantitative outcomes analyzed in this article provide an important overview of the impact of the AF4Q initiative on the Triple Aim, these findings should, by no means, be interpreted as a conclusive statement on the program. A comprehensive assessment of the AF4Q initiative, using both quantitative and qualitative data from all programmatic areas, is provided by Scanlon et al in this supplement.³ Equally important, building local multi-stakeholder health alliances in the participant communities was, by itself, a central goal of the AF4Q program and a potential key to success in achieving the Triple Aim in the long run.⁷ The extent to which the program succeeded in establishing and sustaining such alliances, and their function as neutral conveners in local communities, is discussed by Alexander et al in this supplement.¹³

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 Healthcare Studies, Northwestern University, Feinberg School of Medicine, Chicago, IL (RK, MM); Center for Health Care and Policy Research, Penn State University, University Park, PA (YM, DPS); 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, Mr Kang, Ms Mahmud, Dr McHugh, Dr Scanlon, and Dr Shi and report receipt of grants from RWJF. Dr Greene reports meeting or conference attendance on behalf of Insignia Health. Dr Scanlon reports receipt of grants from RWJF and meeting or conference attendance on behalf of RWJF.

Authorship information: Concept and design (JAA, JBC, JG, MJJ, RK, YM, MM, DPS, YS); acquisition of data (JBC, DPS, YS); analysis and interpretation of data (JAA, JBC, JG, MJJ, RK, YM, MM, DPS, YS); drafting of the manuscript (JG, RK, YM, DPS, YS); critical revision of the manuscript for important intellectual content (JAA, JG, MJJ, RK, MM, DPS, YS); statistical analysis (RK, DPS, YS); obtaining funding (DPS); and supervision (DPS, YS).

Address correspondence to: yus16@psu.edu.

REFERENCES

1. 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.
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(suppl 6):S115-S125.
3. Scanlon DP, Wolf LJ, Alexander JA, et al. Evaluating a complex, multi-site, community-based program to improve health-care quality: the summative research design for the Aligning Forces for Quality initiative. *Am J Manag Care*. 2016;22(suppl 12):eS8-eS16.
4. Scanlon DP, Shi Y, Bhandari N, Christianson JB. Are health-care quality "report cards" reaching consumers? Awareness in the chronically ill population. *Am J Manag Care*. 2015;21(3):236-244.
5. Hibbard JH, Greene J, Shi Y, Mittler J, Scanlon D. Taking the long view: how well do patient activation scores predict outcomes four years later? *Med Care Res Rev*. 2015;72(3):324-337. doi: 10.1177/1077558715573871.
6. Greene J, Fuentes-Caceres V, Verevkina N, Shi Y. Who's aware of and using public reports of provider quality? *J Health Care Poor Underserved*. 2015;26(3):873-888. doi: 10.1353/hpu.2015.0093.
7. 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.
8. Cebul RD, Dade SE, Letourneau LM, Glaseroff A. Regional health improvement collaboratives needed now more than ever: program directors' perspectives. *Am J Manag Care*. 2012;18(suppl 6):S112-S114.
9. 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.
10. 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.
11. 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.
12. 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.
13. 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.
14. James J. Findings of the Oregon Health Insurance Experiment. www.rwjf.org/en/library/research/2015/07/the-oregon-health-insurance-experiment.html. Published July 2015. Accessed July 21, 2016.
15. McHugh M, Harvey JB, Kang R, Shi Y, Scanlon DP. Measuring the dose of quality improvement initiatives. *Med Care Res Rev*. 2016;73(2):227-246. doi: 10.1177/1077558715603567.
16. McHugh M, Shi Y, Ramsay PP, et al. Patient-centered medical home adoption: results from Aligning Forces for Quality. *Health Aff (Millwood)*. 2016;35(1):141-149. doi: 10.1377/hlthaff.2015.0495.
17. The Dartmouth Atlas of Health Care website. www.dartmouthatlas.org/. Accessed July 21, 2016.
18. MarketScan research databases. Truven Health Analytics website. <http://truvenhealth.com/your-healthcare-focus/analytic-research/marketscan-research-databases>. Accessed July 21, 2016.
19. Behavioral Risk Factor Surveillance System. CDC website.

www.cdc.gov/brfss/. Accessed July 21, 2016.

20. CAHPS Hospital Survey. Hospital Consumer Assessment of Healthcare Providers and Systems website. www.hcahpsonline.org/home.aspx. Accessed July 21, 2016.

21. Bertrand M, Duflo E, Mullainathan S. How much should we trust differences-in-differences estimates? National Bureau of Economic Research website. www.nber.org/papers/w8841.pdf. Working Paper 8841. Published March 2002. Accessed July 21, 2016.

22. Ai C, Norton EC. Interaction terms in logit and probit models. *Econ Lett*. 2003;80(1):123-129. doi:10.1016/S0165-1765(03)00032-6.

23. Gruber J. The incidence of mandated maternity benefits. *Am Econ Rev*. 1994;84(3):622-641.

24. Rickles J. A review of *Propensity Score Analysis: Fundamentals and Developments*. *J Educ Behav Stat*. 2016;41(1):109-114. doi: 10.3102/1076998615621303.

25. Abadie A, Diamond A, Hainmueller J. Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *J Am Stat Assoc*. 2010;490(105):493-505. doi: 10.1198/jasa.2009.ap08746.

26. McHugh M, Shi Y, McClellan SR, et al. Using multi-stakehold-

er alliances to accelerate the adoption of health information technology by physician practices. *Healthc (Amst)*. 2016;4(2):86-91. doi: 10.1016/j.hjdsi.2016.01.004.

27. Bauchner H, Fontanarosa PB. The future of US health care policy. *JAMA*. 2016;315(13):1339-1340. doi: 10.1001/jama.2016.2447.

28. Schneider AL, Kalyani RR, Golden S, et al. Diabetes and pre-diabetes and risk of hospitalization: the Atherosclerosis Risk in Communities (ARIC) Study. *Diabetes Care*. 2016;39(5):772-779. doi: 10.2337/dc15-1335.

29. Brennan JJ, Chan TC, Killeen JP, Castillo EM. Inpatient readmissions and emergency department visits within 30 days of a hospital admission. *West J Emerg Med*. 2015;16(7):1025-1029. doi: 10.5811/westjem.2015.8.26157.

30. McHugh M, Harvey JB, Kang R, Shi Y, Scanlon DP. Community-level quality improvement and the patient experience for chronic illness care. *Health Serv Res*. 2016;51(1)76-97. doi: 10.1111/1475-6773.12315.

30. McHugh M, Harvey JB, Kang R, Shi Y, Scanlon DP. Community-level quality improvement and the patient experience for chronic illness care. *Health Serv Res*. 2016;51(1)76-97. doi: 10.1111/1475-6773.12315.