Research Report

Evaluation of CMS's Federally Qualified Health Center (FQHC) Advanced Primary Care Practice (APCP) Demonstration

Final Second Annual Report

Katherine L. Kahn, Justin W. Timbie, Mark W. Friedberg, Tara A. Lavelle, Peter Mendel, J. Scott Ashwood, Liisa Hiatt, Ian Brantley, Beverly A. Weidmer, Afshin Rastegar, Aaron Kofner, Rosalie Malsberger, Mallika Kommareddi, Denise D. Quigley, Claude Messan Setodji

July 2015

RR-886/1-CMS

Prepared for the Centers for Medicare & Medicaid Services

For more information on this publication, visit www.rand.org/t/rr886z1

Published by the RAND Corporation, Santa Monica, Calif.

Table of Contents

Exhibits

Executive Summary

In December 2009, President Barack Obama directed the Department of Health and Human Services, acting through the Centers for Medicare and Medicaid Services (CMS), to implement a three-year demonstration intended to support the transformation of federally qualified health centers (FQHCs) into advanced primary care practices (APCPs) in support of Medicare beneficiaries. FQHCs serve an important function nationally as organizations funded by Section 330 of the Public Health Services Act to offer primary health care to underserved populations. APCPs are designed to encourage doctors, hospitals, and other health care providers to work together to better coordinate care for patients. The demonstration supported the transformation of FQHCs through the use of APCP principles that are designed to support continuous, comprehensive, patient-centered medical care.

For the demonstration, CMS recognizes as advanced primary care (APC) the type of care that is offered by FQHCs that have achieved Level 3 recognition as a patient-centered medical home (PCMH) from the National Committee for Quality Assurance (NCQA). PCMH recognition is based on scoring according to six standards (enhancing access and continuity, identifying and managing patient populations, planning and managing care, providing self-care support and community resources, tracking and coordinating care, and measuring and improving performance), each of which comprises multiple elements. Sites achieve Level 1, 2, or 3 recognition based on their total number of points scored across elements. An expectation of the demonstration is that all FQHCs receiving the demonstration's interventions would attain Level 3 within three years.

The goals of the demonstration are to improve the safety, effectiveness, efficiency, timeliness, and quality of care; patient access to care; adherence to evidence-based guidelines; care coordination and care management; and patient experiences with care. These improvements, in turn, may lead to better health outcomes and management of chronic conditions, decreased use of certain health care services (e.g., hospitalizations, emergency department [ED] visits, duplicative/unnecessary tests and procedures), increased use of other services (e.g., preventive services), and reductions in health care expenditures.

Demonstration Components

To achieve these goals, support FQHCs' transformation into PCMHs, and provide APC to Medicare fee-for-service beneficiaries, the demonstration provides four intervention components to the 500 selected demonstration sites. These components, delivered by a network of organizations from November 2011 through October 2014, are as follows:

1. CMS provides participating FQHCs with a quarterly care management payment of $18 for each eligible Medicare beneficiary.
2. The NCQA offers technical assistance (TA) to help participating FQHCs obtain Level 3 PCMH recognition (based on 2011 NCQA standards). Among other things, participating FQHCs are offered assistance to help them complete semiannual Readiness Assessment Surveys (RASs) and prepare documentation for NCQA PCMH recognition.
3. Through an extensive learning system involving the Health Resources and Services Administration, CMS subcontractor American Institutes for Research (AIR), and primary care associations (PCAs), FQHCs receive training and assistance to support and guide them in their transformation into APCPs.
4. Participating FQHCs periodically receive feedback reports. The first and second reports are at the FQHC level; the third includes beneficiary-level data. The first two allow FQHCs to track their performance on the RASs and to compare their performance with other demonstration sites. The third tracks FQHC performance on key cost and utilization measures for Medicare beneficiaries attributed to that FQHC.

Evaluation

RAND is conducting an independent evaluation of the FQHC APCP demonstration for CMS. The evaluation includes studying the processes and challenges involved in transforming FQHCs into APCPs and assessing the effects of the APCP model on access, quality, and cost of care provided to Medicare and Medicaid beneficiaries currently served by FQHCs. The evaluation measured these trends for FQHCs participating in the demonstration, relative to trends in the same measures for a comparison group of nonparticipating FQHCs.

RAND is providing two interim reports and a third, final report on the demonstration evaluation. This report is the second in the series and presents methodological advances, successful evolution of our data sets, and information describing the demonstration and

the characteristics of the sites, clinicians, staff, administrators, and beneficiaries associated with those sites, as well as the description of the FQHC comparison sites. It addresses three key policy questions. Briefly, Key Policy Question 1 asks about the effects of the demonstration on NCQA recognition and other measures of practice change. Key Policy Question 2 asks whether demonstration sites deliver better beneficiary processes and outcomes than comparison sites. Key Policy Question 3 asks which practice-site and beneficiary characteristics are associated with observed changes in structures, processes, and outcomes.

Conceptual Model

RAND will use Donabedian's classic quality-of-care model to anchor the key study questions.[1] In this model, *structure* conveys the attributes of the settings in which health care occurs. Structure includes material resources (facilities, equipment, and funding) and human resources, including practice organization, quality review, and reimbursement methods. *Process* describes services provided for patients related to diagnostics or therapeutics. *Outcomes* indicate what happens to patients as defined by the effects of care on health status for patients and populations. Donabedian's model specifies that good structure increases the realization of good process, which then increases the realization of valued outcomes. Consistent with this model, we conceptualize that the interventions associated with CMS's FQHC APCP demonstration are designed to first affect the structures of FQHCs. We hypothesize that interventions to transform FQHCs to APCPs can activate the quality-of-care cascade with resultant changes in structures, processes, and outcomes. Exhibit S.1 shows the building blocks of Donabedian's model and how they map to the evaluation of the FQHC APCP demonstration.

[1] Donabedian, A. (1980). The definition of quality and approaches to its assessment. *Explorations in quality assessment and monitoring, Vol. 1*. Ann Arbor, Mich.: Health Administration Press; Donabedian, A. (1982). The criteria and standards of quality. *Explorations in quality assessment and monitoring, Vol. 2*. Ann Arbor, Mich.: Health Administration Press; Donabedian, A. (1988). The quality of care: How can it be assessed? *JAMA: The Journal of the American Medical Association, 260*(12), 1743–1748.

Exhibit S.1: Building Blocks of Donabedian's Model Map to the Evaluation of the FQHC APCP Demonstration

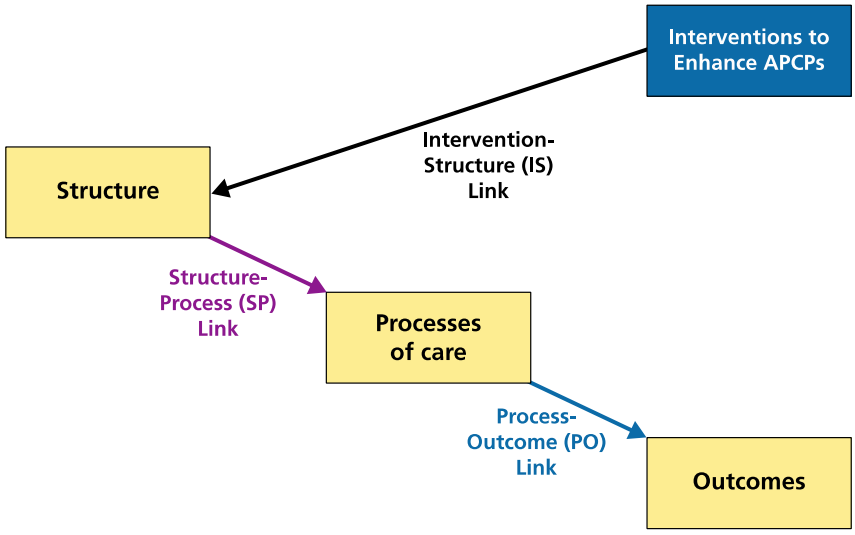


Exhibit S.1 highlights the conceptual model that underlies the demonstration and its corresponding evaluation. CMS's demonstration interventions are designed to improve primary care practice attributes. The evaluation will measure the extent to which exposure to the demonstration is associated with changes in structures as measured by NCQA Level 3 recognition. Furthermore, it will assess how exposure to the demonstration and changes in structures are associated with changes in processes and outcomes of care.

The evaluation translates this conceptual framework into the three key policy questions mentioned above. Our findings in regard to each of these questions are described next.

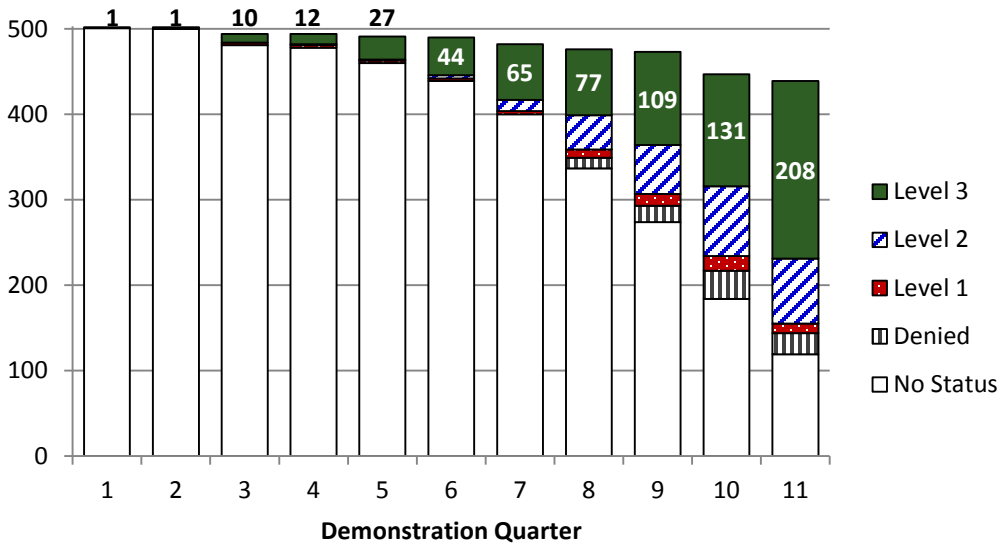
Key Policy Question 1: What Are the Effects of Interventions Designed to Stimulate APC Principles on Changes in Practice Structure and on NCQA Recognition Status?

The first key policy question focuses on FQHC achievement of Level 3 NCQA status. Our analyses address demonstration site progress in achieving NCQA Level 3 recognition and advancing RAS scores; the extent to which demonstration FQHCs utilize the intervention components; factors that predict FQHC use of the demonstration interventions, including TA; and the process of transformation.

CMS set a goal of NCQA Level 3 recognition for 90 percent of the demonstration FQHCs. Our analyses show that demonstration FQHCs had a slow start toward achieving

Level 3 recognition but that the pace accelerated toward the end of the demonstration. By August 2014, 208 of the participating 439 FQHCs (47 percent) had achieved Level 3 PCMH recognition, 76 (17 percent) had achieved Level 2 recognition, and 11 (3 percent) had achieved Level 1 recognition (see Exhibit S.2). Of these 439 FQHCs participating in August 2014, 144 sites (33 percent) had no recognition; of those, 25 (6 percent of the total 439) were denials and 119 (27 percent) either never applied or were still awaiting results. Over time, with the maturing experience of the teams delivering TA to the demonstration sites and the awareness that time to achieve Level 3 recognition within the course of the demonstration was dwindling, the number of recognized FQHCs has been increasing. As the demonstration progressed, we observed a substantial increase in the number of recognized sites from the preceding quarter, when only 131 participating FQHCs (28 percent) had achieved Level 3 recognition. By October 2014, 55 percent of demonstration FQHCs had achieved NCQA Level 3 recognition, compared with 11 percent of comparison FQHCs.[2]

Exhibit S.2: Trends in NCQA Recognition Levels Achieved by Participating FQHCs, Demonstration Quarters 1–11 (N=439)



SOURCE: RAND analysis of CMS NCQA PCMH Recognition Status Data (current as of August 20, 2014).
NOTE: Sites labeled "No Status" include both sites that have not applied for NCQA recognition and those that have applied but have yet to receive a score.

[2] The bulk of this report shows data through August 20, 2014, but for the comparison of rates of NCQA achievement for demonstration versus comparison FQHCs, we have been able to supplement these reports with updated data from HRSA through October 2014.

Trends in self-reported RASs have also been upward, with particularly dramatic improvement in planning and managing care, and lesser but significant advances in measuring and improving performance. Current RAS point scores range between 75 percent and 90 percent across the six PCMH standards, up from 60 percent or less for most of them as of two and a half years ago.

We examined variability in FQHC usage of the four components of the intervention. Median payments to FQHCs, a function of the number of Medicare beneficiaries attributed to each FQHC, were $6,500 per quarter, but varied greatly across FQHCs in the same quarter. Many site leaders said in interviews that they did not know how the payments were used, but remarked that transformation costs spanned multiple domains and years, while the payments accrued only during demonstration years, only to participating FQHCs, and only per Medicare patient.

While TA associated with NCQA and AIR was initiated later and at lower rates than anticipated, TA delivery was more standardized by the end of the second demonstration year. Though TA participation rates showed substantial regional variations throughout the demonstration, efforts by PCAs working in conjunction with AIR were apparent in all regions of the nation, with PCAs providing TA to sites by leading conference calls, webinars, site visits, email exchanges, and calls. Despite regional differences in usage of site-level feedback reports designed by CMS to enhance site-level awareness of utilization by their Medicare beneficiaries, 68 to 92 percent of the FQHCs across all regions logged in to download feedback data by the end of the demonstration.

Given the observed variability in use of the demonstration components by the FQHCs, we examined what factors predict higher (or lower) use. Our assessment of predictors of TA use found higher TA use was associated with more baseline FQHC resources, such as PCMH payments from one or more plans and total site revenue. In contrast, high-baseline RASs and ambulatory care quality accreditation were associated with lower TA use, perhaps because of a lower dependency on TA for progress to PCMH recognition.

We also evaluated how TA supports transformation and have drawn preliminary conclusions. (They are preliminary because they use RAS scores as a proxy for transformation while we await the Level 3 recognition scores from NCQA, which are expected some months after the completion of the demonstration.) We did not find the statistically significant positive associations that we expected between TA usage and high RAS scores. We did find that a site's participation in a prior CMS shared-savings demonstration (most notably CMS's Shared Savings Program) was one strong site-level predictor of achieving a Level 3 score on the 30-month RAS. Sites that participate in shared-savings demonstrations may benefit from additional resources gained from

participating in these demonstrations, which facilitate achievement of Level 3 recognition. In addition, the potential for shared savings may serve as a potent driving force behind these sites' pursuit of Level 3 recognition.

Key Policy Question 2: Do Demonstration Sites Deliver Better Beneficiary Processes and Outcomes than Comparison Sites?

Key Policy Question 2 asks whether Medicare beneficiaries assigned to demonstration sites experience better processes and outcomes of care than those experienced by beneficiaries at comparison sites. To assess the effectiveness of the APCP demonstration, Key Policy Question 2 focuses on differences between demonstration and comparison sites that may lead to better management of chronic conditions, improved health outcomes, decreased use of inappropriate health care services (e.g., certain hospitalizations, ED visits, duplicative or unnecessary tests and procedures), increased use of other appropriate services (e.g., preventive services), and reductions in health care expenditures.

To answer this policy question, the evaluation focuses on metrics spanning 12 research questions (RQs) pertinent to APCP principles, including: (1) continuity, (2) timeliness, (3) access to care, (4) adherence to evidence-based guidelines, (5) beneficiary ratings of providers, (6) effective beneficiary participation in decision making, (7) self-management, (8) patient experiences with care, (9) coordination of care, (10) disparities, (11) utilization, and (12) expenditures. Some of these metrics are evaluated using claims data, others by beneficiary report. Both data sources are needed to meaningfully understand whether demonstration interventions designed to stimulate APCP principles are associated with improvements in processes and outcomes of care for Medicare beneficiaries. Claims analyses provide the advantage of a very large sample with a longitudinal data analysis; beneficiary survey analyses provide rich clinical and patient experiential data. Also included in our evaluation of Key Policy Question 2 are data from the clinician and staff experience (CASE) survey and qualitative analyses that provide context and nuance for interpreting all the available data.[3]

We have thus collected and analyzed data from multiple sources that connect the experiences of patients, providers, staff, and policymakers with the demonstration and the transformation of care toward APC practices. Claims-based analyses pertinent to

[3] By design, these follow-up primary data collection efforts were planned toward the end of the demonstration so that our longitudinal analyses could maximize the difference between early (baseline) and late (follow-up) experiences. The final report will include longitudinal analyses of all three of these primary data sources as a supplement to a more complete set of claims data.

utilization, cost, evidence-based care, and continuity of care are summarized here, followed by a summary of beneficiary survey analyses. After selecting evaluation metrics for each of these RQs, we assessed the impact of the demonstration on these metrics by comparing the FQHC demonstration cohort with comparison cohorts.

As of the end of the demonstration's ninth quarter, we know that costs for demonstration sites are higher than for comparison sites. For example, during the last three quarters of the demonstration analyzed to date, the per-person-per-quarter total cost for each Medicare beneficiary is $65–101 more in demonstration sites than in comparison sites relative to differences at baseline. These cost findings are important and suggest that it is unlikely that overall costs associated with the demonstration at is completion will be lower for demonstration FQHCs than for comparison FQHCs.

Claims-Based Utilization and Cost Findings

Across nine quarters comparing demonstration FQHCs with comparison FQHCs adjusted for differences at baseline, we noted significantly more utilization for three of four claims-based utilization measures,[4] and more total costs for the demonstration FQHC users than for the comparison FQHC users. In contrast with comparison FQHC users, demonstration FQHC users had significantly more

- total Medicare payments in four of nine quarters
- hospital admissions in two of nine demonstration quarters
- admissions for ambulatory care sensitive conditions (ACSCs) in none of nine demonstration quarters
- unplanned hospital readmissions in one of nine demonstration quarters, and
- ED visits in six of nine demonstration quarters.

There was no demonstration quarter in which demonstration site users had significantly less utilization or less total cost than comparison sites.

Claims-Based Evidence-Based Findings

Across nine quarters comparing demonstration FQHCs with comparison FQHCs adjusted for differences at baseline, the demonstration FQHC group significantly outperformed the comparison FQHC group for at least eight quarters for glycated hemoglobin blood tests (HbA1C), retinal eye exams, and nephropathy testing, though we

[4] Significant differences are defined as $p<0.05$ for differences between demonstration and comparison site relative to differences at baseline.

observed little difference for low-density lipoprotein (LDL) testing for diabetic patients or for lipid testing for ischemic vascular disease patients.

Claims-Based Continuity of Care Findings

Adjusted analyses show no statistically significant differences in the Bice-Boxerman continuity of care measure between the demonstration and comparison FQHC groups in the first year of the demonstration compared with baseline. In year 2, demonstration FQHCs are associated with a statistically significant but very small (~1 percent) decrease in continuity when looking across all primary care provider visits and when looking at primary care and specialist care together.

Beneficiary Survey-Based Findings

Of the 46 beneficiary survey–based metrics, 85 percent of the primary analyses comparing demonstration FQHCs with comparison FQHCs showed no significant differences at this stage of the analysis. Further beneficiary survey analyses will be included in the final report after longitudinal data become available.

Summary of Differences Between Demonstration and Comparison Findings to Date

In summary, relative to four baseline quarters, claims-based analyses across nine quarters show significantly more utilization and costs for demonstration FQHCs than comparison FQHCs for total Medicare payments (four quarters), hospital admissions (two quarters), readmissions (one quarter), and ED visits (six quarters). Standard claims-based process measures characterizing utilization among Medicare beneficiaries with diabetes or ischemic vascular disease show improvements in process for demonstration FQHC users compared with comparison FQHC users after considering baseline differences. Most beneficiary survey analyses show no difference in cross-sectional analyses, but longitudinal beneficiary analyses will not be available until the final report.

Our work during the next year will include the completion of analyses of recently and yet-to-be-collected data paired with the application of new methods, which will allow us to assess whether the demonstration was associated with differences in beneficiary experiences, and if so, whether those experiences were associated with changes in beneficiary-reported outcomes. At this time, we do not yet know whether the observed additional costs for demonstration sites are associated with better (or worse) clinical outcomes for FQHC users. That will happen at the end of the demonstration, when we

have additional longitudinal beneficiary, clinician and staff, and claims data with contextual qualitative analyses that will be critical to our completing this evaluation.

Key Policy Question 3: Which Practice-Site and Beneficiary Characteristics Are Associated with Observed Changes in Structures, Processes, and Outcomes?

Key Policy Question 1 informs us about predictors of TA uptake and how TA and other features contribute to a site's transformation to Level 3. Key Policy Question 2 examines whether demonstration sites deliver better beneficiary processes and outcomes than comparison sites. Analyses from these two questions will inform our final policy question. Key Policy Question 3 asks which practice-site and beneficiary characteristics are associated with observed changes in structures, processes, and outcomes. This question will be answered during the next and final period of the evaluation. The finding of any statistically significant and substantive demonstration impact raises the question of whether the changes in beneficiary outcomes are due to the implementation of APCPs or other causes. Answering this question is critically important not only for appropriately interpreting the results of the FQHC APCP demonstration, but also for formulating sound policies and programs based on the evaluation's findings. We approach Key Policy Question 3 with three methodological policy questions that allow us to extend our understanding of the demonstration and the mechanisms by which beneficiary outcomes are affected through a site's participation in the demonstration and a site's achievement of Level 3 recognition. These are as follows:

- Key Policy Question 3.1A: Is the impact of the demonstration on beneficiary outcomes mediated by the adoption of APCPs?
- Key Policy Question 3.1B: To what extent do demonstration sites improve beneficiary outcomes in ways that are not explained by achievement of Level 3 PCMH recognition?
- Key Policy Question 3.1C: What is the impact of achieving Level 3 PCMH recognition on beneficiary outcomes independent of a site's participation in the demonstration?

The analyses supporting these questions account for the measured changes in practice structure that accompany the transition to a PCMH and will help to differentially estimate the impact on beneficiary outcomes from participation in the demonstration that is not explained by Level 3 recognition and that is explained by PCMH recognition independent of participation in the demonstration. During the remainder of the evaluation, we will

address these questions using a unified modeling framework that estimates both impacts simultaneously, using a difference-in-differences framework similar to the overall evaluation framework but with additional information specifying the PCMH recognition status of each site. Only sites that have achieved NCQA Level 3 recognition will be designated as PCMH-recognized sites. These analyses will compare how demonstration and comparison groups differ in the last year of the demonstration compared with their differences in the baseline year, and how PCMH-recognition sites measure up against nonrecognized sites in the last year compared with the baseline year.

Building upon this analysis, Key Policy Question 3 also examines the extent to which the demonstration's impact on beneficiary outcomes is stronger for certain types of sites and beneficiaries as compared with other types. We will examine this phenomenon, also known as *effect modification*, across a range of characteristics hypothesized to influence the magnitude of the demonstration's impact, including:

- beneficiary characteristics: dual eligibility, disability, comorbidity
- site characteristics: number of beneficiaries, total revenue per site, number of affiliated service delivery sites within the grantee's organization
- area-level characteristics: percentage of poverty in the census tract in which the FQHC operates, location in a rural area.

We will also examine effect modification by a site's baseline readiness for becoming a medical home, which will be measured by baseline RAS scores. Analyses using baseline RAS scores, unlike all other analyses in this section, will be limited to the subset of demonstration and comparison sites for which we have baseline RAS data.

The next set of analyses pertinent to Key Policy Question 3 seeks to better understand the demonstration impacts we observed in Key Policy Question 2. These analyses comprise two types. First, we identify the factors that enable demonstration sites to achieve higher levels of performance relative to comparison sites on each of several key intermediate outcomes (e.g., timeliness of care, beneficiary experience, delivery of evidence-based care). Second, we assess the extent to which changes in these intermediate outcomes explain the observed patterns in cost and utilization outcomes for demonstration sites relative to comparison sites over the course of the demonstration. Taken together, these analyses provide additional context for interpreting the demonstration's impact.

Supplementing these analyses, Key Policy Question 3 assesses differences and commonalities among the high- and low-performing FQHCs and offers suggestions to CMS for making changes if the demonstration were to be extended or expanded. Here, we review feedback and suggestions for future improvements in the demonstration and PCMH initiatives from three qualitative sources. At site leader interviews conducted at the

end of the baseline, we asked respondents whether they had any questions, concerns, or feedback to bring to the attention of CMS and its partners who have organized the initiative. We also asked baseline PCA leader interview respondents if they had recommendations for feedback; again, they suggested a few changes for the demonstration and future initiatives. We supplement these reports with recommendations for improving future PCMH TA based on our previous qualitative analysis of perspectives on TA intervention components from the baseline site interviews.

These suggestions include:

- providing better coordination among a more limited number of TA providers
- examining better adaptation of the NCQA recognition standards and process to FQHCs, and other changes in support for preparing the NCQA application
- ensuring sustainability of PCMH transformation and recognition, including financial support
- increasing the time for sites to lay the groundwork and pilot PCMH changes
- starting in-person training and direct assistance as early as possible in the demonstration
- focusing and formatting feedback reports to provide actionable information for sites
- making changes to support PCAs in their roles as TA providers.

Conclusion

Summary of the Second Annual Report

In sum, the APCP demonstration supported by CMS is an enormous undertaking. The three-year demonstration aims to support FQHCs and the beneficiaries they serve by enhancing efforts to achieve APC. Participating FQHC sites are expected to obtain NCQA Level 3 PCMH recognition within the span of the three-year demonstration, motivated and supported by four components of the intervention.

How is progress toward that goal? Of the three key policy questions addressed by RAND's evaluation team, the first pertains to the impact of the demonstration components on achievement of NCQA Level 3 recognition. Though sites had a slow start in achieving NCQA Level 3 recognition, significantly more demonstration than comparison FQHCs (69 versus 11 percent) achieved NCQA Level 3 recognition by October 2014. Reflecting the diversity of demonstration sites with respect to their baseline TA needs, the approach toward TA evolved as the demonstration progressed, with persistent planning, iterative efforts to respond to noted challenges, and extensive implementation protocols. During the

latter half of the demonstration's second year, TA was delivered more systematically and the collection of stable metrics for assessing TA was implemented. With time, TA efforts became more individualized to site-specific needs, more sites participated in TA activities, and more sites accessed feedback reports. Still, many sites have reported preferring to wait as long as possible to prepare the strongest Level 3 recognition application they can. This has presented problems for the evaluation, because Level 3 recognition has not been a reliable means for distinguishing the transformation of FQHCs to APCPs. That is to be resolved when NCQA makes its recognition decisions (expected January 2015).

Our second key policy question asks whether demonstration sites deliver better processes and outcomes than comparison sites do. We ask this question focusing on a diverse set of clinically meaningful processes, as well as outcomes including utilization, cost, and use of evidence-based care. Our preliminary analyses to date show significant differences between demonstration and comparison FQHCs for most measured claims-based metrics, but no differences for most beneficiary survey–based metrics. Claims-based analyses across nine quarters show significantly more utilization and costs for demonstration FQHCs compared with comparison FQHCs for total Medicare payments (four quarters), hospital admissions (two quarters), readmissions (one quarter), and ED visits (six quarters). As of the writing of this report, it appears that the demonstration will not achieve cost savings when compared with comparison FQHC sites.

Our approach to the third key policy question—which asks which practice-site and beneficiary characteristics are associated with changes in structures, processes, and outcomes—is pending access to final longitudinal data that includes the demonstration's final quarters. When these data are available, we will apply an integrated analysis plan to the assessment of participation in the demonstration and achievement of NCQA Level 3 impact beneficiary outcomes.

While we have claims results for nine quarters that show that the demonstration is not likely to achieve cost savings when compared with comparison FQHC sites, final assessment of the demonstration impact will not be known until final claims analyses are supplemented with longitudinal beneficiary analyses. These analyses will be reported within the final report.

A substantial body of literature has emerged in recent years regarding the effectiveness of PCMH transformation; many results have been mixed.[5] Hoff et al. reviewed 21 studies related to PCMHs from 2007 through 2010.[6] Seven of these studies measured quality of

[5] Peikes, D., Zutshi, A., Genevro, J. L., Parchman, M. L., and Meyers, D. S. (2012). Early evaluations of the medical home: Building on a promising start. *American Journal of Managed Care 18,* 105–116.

[6] Hoff, T., Weller, W., and Depuccio, M. (2012). The patient-centered medical home: A review of recent research. *Medical Care Research and Review, 69,* 619–644.

care as an outcome, and all seven demonstrated that, compared with other practices, PCMHs were associated with better quality. But studies that measured ED utilization, inpatient utilization, expenditures, and beneficiary experience in PCMHs were less convincing. Peikes and Chen found mixed evidence of impacts when they analyzed six studies with rigorous quantitative analyses regarding health, health care, and costs.[7] Jackson et al. noted 19 comparative studies demonstrating small to moderate positive effects on patient and staff experiences and delivery of preventive services, as well as reductions in ED visits, but no cost savings or reductions in hospital admissions in older adults.[8]

Furthermore, in January 2014, the Patient-Centered Primary Care Collaborative published its annual update on the evidence of the PCMH's impact on cost and quality.[9] The report focused on 20 studies, including 13 peer-reviewed/academic and seven industry-reported studies released between August 2013 and December 2013. Their report indicated cost reductions for 61 percent of peer-reviewed studies and 57 percent of industry reported studies. Respectively, they note 61 percent of peer-review and 57 percent of industry studies report fewer ED visits; 31 percent of peer-review and 57 percent of industry studies report fewer inpatient admissions, and 13 percent of peer-reviewed and 29 percent of industry studies reported fewer readmissions. For adherence to evidence-based measures, access, and satisfaction, less than one-third of the peer-reviewed studies and a smaller percentage of industry studies reported improvements with the implementation of PCMH initiatives. The authors cite these findings as early, and as not restricted to a formal peer-review meta-analysis of the literature. Nevertheless, they conclude that many studies show that fully transformed primary care practices demonstrated desired outcomes, and that the diversity of designs across many studies, as expected, shows a spectrum of results.

Achievement of NCQA recognition status has been cited by many as a valid measure of a practice's adoption of APCP principles. This status is expected to translate into

---

[7] Peikes, D., and Chen, A. (2009). Effects of care coordination on hospitalization, quality of care, and health care expenditures among Medicare beneficiaries. *Journal of the American Medical Association, 301*(6): 603–618.

[8] Jackson, G. L., Powers, B. J., Chatterjee, R., Prvu Bettger, J., Kemper, A. R., Hasselblad, V., Dolor, R. J., Irvine, J., Heidenfelder, B. L., Kendrick, A. S., Gray, R., and Williams, J. W. (2013). The patient-centered medical home: A systematic review. *Annals of Internal Medicine, 158,* 169–178.

[9] Nielsen, M., Olayiwola, J. N., Grundy, P., and Grumbach, K. (2014). *The patient-centered medical home's impact on cost and quality: An annual update of the evidence, 2012–2013*. Washington, D.C.: Patient-Centered Primary Care Collaborative; 2014. Accessed October 1, 2014: http://www.milbank.org/uploads/documents/reports/Patient-Centered_Medical_Homes_Impact_on_Cost_and_Quality.pdf.

improved performance on CMS's triple aims of improving health, improving health care, and lowering costs.[10] In fact, despite substantial growth in the number of practices pursuing APC attributes and the support being offered to them, evidence supporting the effectiveness of PCMH transformation in improving both quality and cost/utilization outcomes has been mixed. The results from studies evaluating the effectiveness of medical home recognition have been associated with a growing interest in what makes an effective medical home. It will take the analysis of additional data and the application of additional methods for the evaluation team to distinguish what components of the demonstration and characteristics of participating sites in the demonstration are associated with improved processes and outcomes. These analyses are planned during this next year. Many are described in the discussion of Key Policy Question 3.

What Can We Expect from the Final Report?

At the time of the writing of this report, the demonstration is coming to a close and the many stakeholders who have participated in and been affected by the demonstration are sharing their stories. Through site visits, longitudinal site leader interviews, PCA interviews, PCA focus groups, and beneficiary and caregiver focus groups, we have been learning how clinics aspire to and often do achieve NCQA recognition. We have recently completed our fielding of the final CASE survey and received our final RAS data. During the next months we will access our final quarters of claims data, and complete our final cohort of late (follow-up) beneficiary surveys. We anticipate receiving the final demonstration-associated NCQA recognition data for demonstration and comparison sites. With the influx of all of these data sets, we will, for the first time, have available longitudinal data across multiple modalities. As planned, the final year of our analysis will involve completing our longitudinal analyses and focusing on our linkage analyses between our different primary and secondary data sources.

While this second annual report presents research findings that have been informed by more than one type of data, the final report will synthesize these data sources, allowing us to tell the evaluation story as planned. Since we will have complete NCQA Level 3 recognition data, we anticipate Key Policy Question 1 analyses will more clearly articulate the predictors of Level 3 recognition. Our regression methodologies will be supplemented by qualitative comparative analyses, and these will be supplemented by thematic analyses, each exploring why some sites achieve Level 3 recognition early, some late, and some not at all.

---

[10] Berwick, D. M., Nolan, T. W., and Whittington, J. (2008). The triple aim: Care, health, and cost. *Health Affairs, 27*(3), 759–769.

The addition of longitudinal beneficiary data will enhance claims analyses of Key Policy Question 2. For the first time with this evaluation, we will be able to tell the stories of beneficiaries over time with breadth. CASE survey data and RAS/NCQA data will provide important variables to allow us to understand how practices change, as well as the impact of practice change on beneficiaries. Qualitative analyses will enrich our understanding of these data and help us to prioritize our inquiries.

While this report has introduced a conceptual model and tools for understanding predictors of change, having these linked data files will allow us to more fully pursue the RQs of Key Policy Questions 2 and 3. This reports shows that demonstration costs are significantly higher than comparison costs for four demonstration quarters. It is not likely that these demonstration cost trends are likely to reverse by the end of the demonstration. While we examine the remaining quarters of cost data, we will also be examining longitudinal beneficiary survey–report data and plan to report whether beneficiary experiences change as the demonstration is completed. Further, we anticipate being able to distinguish which changes in beneficiary experience relate to the achievement of NCQA Level 3 compared with other possible explanations. All of these new findings will be included in the final report, along with analyses to help us supplement the story of this demonstration with lessons learned that can inform future demonstrations and effective change.

Glossary

Baseline Period: The year prior to demonstration initiation (November 1, 2010, through October 31, 2011).

Comparison FQHCs: FQHCs selected by RAND for comparison to the demonstration FQHCs.

Demonstration FQHCs: All FQHCs ever selected to participate in the FQHC APCP demonstration (including those participating at demonstration initiation and late entrants).

Demonstration Period: The time period between demonstration initiation (November 1, 2011) and the latest reportable date (the demonstration is ongoing through October 31, 2014).

Dropout FQHCs: Demonstration FQHCs that dropped out of the program, including FQHCs that voluntarily discontinued enrollment and FQHCs with enrollment terminated by CMS.

Late Entrant FQHCs: FQHCs selected to participate in the FQHC APCP demonstration after November 1, 2011.

Participating FQHCs: Demonstration FQHCs participating in the demonstration as of August 26, 2013 (and have not dropped out).

Abbreviations

AAAHC	Accreditation Association for Ambulatory Health Care
ACA	Affordable Care Act
ACSC	ambulatory care sensitive conditions
AIR	American Institutes for Research
APC	advanced primary care
APCP	advanced primary care practice
BMI	Body mass index
CAHPS	Consumer Assessment of Healthcare Providers and Systems
CASE	clinician and staff experience
CG-CAHPS	Clinician and Group Consumer Assessment of Healthcare Providers and Systems
CI	confidence interval
CMS	Centers for Medicare and Medicaid Services
ED	emergency department
EHR	electronic health record
FQHC	federally qualified health center
GEE	generalized estimating equations
GLM	generalized linear model
HbA1c	glycated hemoglobin blood test
HCC	hierarchical condition categories
HCCN	Health Center Controlled Network
HHS	U.S. Department of Health and Human Services
HRSA	Health Resources and Services Administration
ITT	intention to treat
LDL	low-density lipoprotein
MEMO	Minimizing Errors/Maximizing Outcomes
MSIS	Medicaid Statistical Information System
NCQA	National Committee for Quality Assurance
NPI	National Provider Identifier
OR	odds ratio
PBPQ	per beneficiary per quarter
PCA	primary care association
PCC	primary care clinic
PCMH	patient-centered medical home
PHQ–4	Patient Health Questionnaire for Depression and Anxiety
QCA	qualitative comparison analysis
RAS	Readiness Assessment Survey
RHC	rural health centers
RQ	research question

SE	standard error
SF-12	12-Item Short-Form Health Survey
SQ	survey question
TA	technical assistance
UDS	Uniform Data System

I. Introduction

I.1. Overview of the Demonstration

In December 2009, President Barack Obama directed the U.S. Department of Health and Human Services (HHS) to implement a three-year demonstration to support federally qualified health centers (FQHCs) in delivering advanced primary care practices (APCP) to Medicare beneficiaries. FQHCs serve an important function nationally as organizations funded by Section 330 of the Public Health Services Act to offer primary health care to underserved populations. APCP principles are designed to encourage doctors, hospitals, and other health care providers to work together to better coordinate care for patients. The demonstration was designed to support the transformation of FQHCs through the use of APCP principles that facilitate continuous, comprehensive, patient-centered medical care.

The Center for Medicare and Medicaid Services (CMS) is carrying out the demonstration for HHS. For the demonstration, CMS recognizes FQHCs as providing APC once they have obtained Level 3 patient-centered medical home (PCMH) recognition from the National Committee for Quality Assurance (NCQA). Recognition is based on scoring according to six standards: enhancing access and continuity, identifying and managing patient populations, planning and managing care, providing self-care support and community resources, tracking and coordinating care, and measuring and improving performance. Each standard is composed of multiple elements, and sites achieve Level 1, 2, or 3 recognition based on their total number of points scored across elements. An expectation of the demonstration is that all FQHCs receiving the demonstration's interventions would attain Level 3 by the end of the demonstration.

On November 1, 2011, CMS initiated the three-year demonstration intended to support the transformation of FQHCs into advanced primary care practices (APCPs) in support of Medicare beneficiaries. CMS is monitoring each participating FQHC's progress toward obtaining Level 3 NCQA PCMH recognition.

The demonstration provides four intervention components to support FQHC transformation into PCMHs and provision of APC to their Medicare fee-for-service beneficiaries for the 500 selected demonstration FQHCs. It was hypothesized that exposure to these interventions would accelerate FQHC achievement of Level 3 recognition. These components, designed by CMS, are delivered by a network of organizations:

1. To support patient-centered medical care and application for NCQA recognition, CMS provides participating FQHCs with a quarterly care management payment of $18 for each eligible Medicare beneficiary.
2. The NCQA offers technical assistance (TA) to help participating FQHCs obtain NCQA Level 3 PCMH recognition (based on 2011 NCQA standards). Specifically, participating FQHCs are offered assistance to help them complete biannual Readiness Assessment Surveys (RASs) and to prepare documentation for NCQA PCMH recognition.
3. Through an extensive learning system involving the Health Resources and Services Administration (HRSA), the American Institutes for Research (AIR), and primary care associations (PCAs), FQHCs receive training and assistance to support and guide them in their transformation across all six NCQA standards into APCPs.
4. Participating FQHCs periodically receive feedback reports from a CMS subcontractor. The first and second reports are at the FQHC level; the third includes beneficiary-level data. The first two allow FQHCs to track their performance on the RASs over time and to compare their performance with other demonstration sites. The third tracks FQHC performance on key cost and utilization measures over time for attributed Medicare beneficiaries.

Ultimately, the goals of the demonstration are to improve the safety, effectiveness, efficiency, timeliness, and quality of care; patient access to care; adherence to evidence-based guidelines; care coordination and care management; and patient experiences with care. These improvements, in turn, may lead to better health outcomes and management of chronic conditions, decreased use of certain health care services (e.g., hospitalizations, emergency department [ED] visits, duplicative/unnecessary tests and procedures), increased use of other services (e.g., preventive services), and reductions in health care expenditures. To determine whether these goals are met, CMS turned to the RAND Corporation to conduct an independent evaluation.

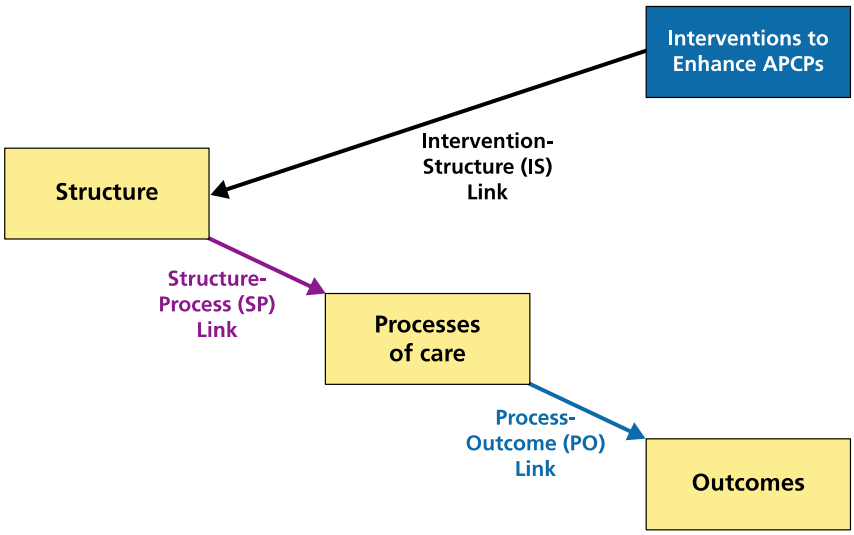
I.2. Overview of the Evaluation

Conceptual Model and Implications

The RAND evaluation addresses the processes and challenges involved in transforming FQHCs into APCPs and assessing the effects of the APCP model on access, quality, and cost of care provided to Medicare and Medicaid beneficiaries served by FQHCs. RAND will use Donabedian's classic quality-of-care model to anchor the

evaluation.[11] In this model, *structure* conveys the attributes of the settings in which health care occurs. Structure includes material resources (facilities, equipment, and funding) and human resources, including practice organization, quality review, and reimbursement methods. *Process* describes services provided for patients related to diagnostics or therapeutics. *Outcomes* indicate what happens to patients as defined by the effects of care on health status for patients and populations. Donabedian's model specifies that good structure increases the realization of good process, which then increases the realization of valued outcomes. Consistent with this model, we conceptualize that the interventions associated with CMS's FQHC APCP demonstration are designed to first affect the structures of FQHCs. We hypothesize that interventions to transform FQHCs to APCPs can activate the quality-of-care cascade with resultant changes in structures, processes, and outcomes. Exhibit I.1 shows the building blocks of Donabedian's model and how they map to the evaluation of the FQHC APCP demonstration.

Exhibit I.1: Building Blocks of Donabedian's Model Map to the Evaluation of the FQHC APCP Demonstration



[11] Donabedian, A. (1980). The definition of quality and approaches to its assessment. In *Explorations in quality assessment and monitoring, Vol. 1*. Ann Arbor, Mich.: Health Administration Press; Donabedian, A. (1982). The criteria and standards of quality. In *Explorations in quality assessment and monitoring, Vol. 2*. Ann Arbor, Mich.: Health Administration Press; Donabedian, A. (1988). The quality of care: How can it be assessed? *JAMA: The Journal of the American Medical Association, 260*(12), 1743–1748.

CMS's demonstration interventions are designed to enhance primary care practice attributes. The evaluation will determine the extent to which exposure to the demonstration is associated with changes in structures as measured by NCQA Level 3 recognition. Furthermore, the evaluation will assess how exposure to the demonstration and changes in structures are associated with changes in processes and outcomes of care, including costs.

Structural changes prompted by the demonstration interventions are expected to facilitate APCP processes for patient populations, with the resulting achievement of improved patient experiences and outcomes. To assess whether patient experiences and outcomes differ for beneficiaries assigned to demonstration sites and those assigned to comparison sites, the evaluation assesses differences across a series of multiple domains. These include continuity of care, timeliness, access, adherence to evidence-based care, beneficiary-rated quality of care, effective participation in decision making, self-management of health problems, experiences with the health care system, utilization of services, expenditures, coordination of care, health care disparities, and costs.

After exploring how exposure to demonstration interventions affects FQHC use of intervention components and achievement of NCQA recognition, as well as how beneficiary experiences at demonstration sites compared with comparison sites, the evaluation will assess whether certain practice-site and beneficiary characteristics are associated with observed changes in structures and beneficiary-level processes and outcomes.

In aggregate, RAND's evaluation will assess which specific practice-site structural attributes are associated with advances in NCQA recognition and which changes in structure are associated with changes in processes and outcomes for beneficiaries and clinics. Furthermore, we will assess predictors of change and how these changes in clinics and with beneficiaries affect such outcomes as utilization and costs. Where data currently allow, all three key policy questions will compare time trends in the FQHC intervention sites with time trends in comparison sites.

Demonstration and Comparison Sites

As indicated above, our main evaluation approach is to evaluate changes in structures, processes, and outcomes for beneficiaries and clinics by assessing demonstration FQHCs against comparison FQHCs. However, we recognize a central challenge with this approach is that efforts to achieve APC and PCMH recognition are prevalent among FQHCs even among those not participating in the demonstration. This could lead to null results from comparing demonstration FQHCs and comparison FQHCs, showing that demonstration participation is not associated with trends in structure, process, and

outcomes. A null finding could reflect either the absence of an effective APCP intervention, or the delivery of a comparably effective intervention to comparison sites during the same time that the APCP intervention is delivered to demonstration sites.

To add to the robustness of this evaluation, we have included plans for sensitivity analyses in which primary care clinics (PCCs) will serve as a secondary comparison group. The most important advantage of a PCC comparison group is that we expect the risk of exposure to PCMH-related transformation support to be lower for PCCs relative to FQHCs. Analyses of PCCs as a comparison group are pending until the PCC data set is complete and ready for analysis. Thus, only demonstration FQHC vs. comparison FQHC analyses are presented in this report.

Key Policy Questions

RAND's proposed evaluation advances three key policy questions that we aim to address by the time of the evaluation's planned completion in September 2016. An analysis for each consecutive policy question depends upon analysis components of the prior question. We begin by studying how the demonstration affects structure and PCMH recognition. We then analyze how the demonstration affects the experiences, processes, and outcomes for beneficiaries attributed to demonstration and comparison groups. Finally, we analyze which practice-site and beneficiary-level characteristics are associated with observed structural and beneficiary level changes.

The evaluation translates the conceptual framework in Exhibit I.1 into three key policy questions, as follows.

I.3A. Key Policy Question 1: What Are the Effects of Interventions Designed to Stimulate APC Principles on Changes in Practice Structure and on NCQA Recognition Status?

The first key policy question focuses on the transformation of demonstration FQHC structures to better support APCP principles as measured by biannual RAS scores and by demonstration FQHCs' achievement of NCQA Level 3 recognition. We also examine predictors of TA uptake and how TA and other features contribute to a site's progression to Level 3. To address this question, RAND has analyzed the four components of the interventions delivered to and utilized by FQHCs participating in the CMS demonstration. These include

- regular management payments to FQHC sites according to the number of attributed Medicare beneficiaries
- access to and use of TA provided by NCQA

- administrative support and TA provided by AIR (CMS's subcontractor), and the regional and state PCAs who work directly with FQHCs to enhance transformation and successful NCQA recognition applications
- access to feedback reports about site- and beneficiary-level performance.

Our analyses address the extent to which demonstration FQHCs utilize the intervention components to integrate new care management strategies, electronic tools, integration of clinical evidence and patient data, and team structures into APCPs expected to improve beneficiary experiences and outcomes and to reduce costs.

I.3B. Key Policy Question 2: Do Demonstration Sites Deliver Better Beneficiary Processes and Outcomes than Comparison Sites?

The second key policy question extends the analysis to effects of the demonstration on beneficiary processes, utilization, and outcomes. Interventions are anticipated to prompt structural changes; these changes are expected to facilitate APC processes for patient populations, with the resultant achievement of improved patient outcomes.

Our design calls for a longitudinal analysis of multiple clusters of beneficiary experiences, processes, and outcomes to determine whether users of demonstration FQHCs experience different care and outcomes than users of comparison FQHCs. We analyze this using both claims and beneficiary survey data.

To better understand how beneficiary experiences change or why they may not, we also analyzed CASE survey and qualitative data. Clinicians, staff, and practice executives provide these inputs that we use to analyze how specific clinic opportunities that were taken and challenges that were experienced might affect beneficiary changes.

I.3C. Key Policy Question 3: Which Practice-Site and Beneficiary Characteristics Are Associated with Observed Changes in Structures, Processes, and Outcomes?

Key Policy Question 3 asks which practice-site and beneficiary characteristics are associated with observed changes in structures, processes, and outcomes. This question will be answered during the final period of the evaluation.

The finding of any statistically significant and substantive demonstration impact raises the question of whether the changes in beneficiary outcomes are due to the implementation of APCPs or other causes. Answering this question is critically important not only for appropriately interpreting the results of the FQHC APCP demonstration, but also for formulating sound policies and programs based on the evaluation's findings.

The analyses supporting this question account for the measured changes in practice structure that accompany the transition to a PCMH. These analyses will help to differentially estimate the impact on beneficiary outcomes from participation in the

demonstration that is not explained by Level 3 recognition and that is explained by PCMH recognition independent of participation in the demonstration. Building upon this analysis, Key Policy Question 3 also examines the extent to which the demonstration's impact on beneficiary outcomes is stronger for certain types of sites and beneficiaries as compared with other types.

The next set of analyses will identify the factors that enable demonstration sites to achieve higher levels of performance relative to comparison sites on each of several key intermediate outcomes (e.g., timeliness of care, beneficiary experience, delivery of evidence-based care) and the extent to which changes in these intermediate outcomes explain the observed patterns in cost and utilization outcomes for demonstration sites relative to comparison sites over the course of the demonstration.

Supplementing these analyses, Key Policy Question 3 assesses commonalities among the high and low-performing FQHCs, giving rise to suggestions to CMS for possible changes if the demonstration were to be extended or expanded.

I.3. Overview of the Second Annual Report

As part of its evaluation strategy, RAND reports to CMS on intermediate evaluation findings each quarter and again annually. At the time of the writing of this second annual report, we have completed 11 quarters of the 12-quarter demonstration, though claims analyses are available only through the end of the ninth quarter and primary data analyses are still in progress. The demonstration is scheduled to end, as planned, October 31, 2014.

This annual report is organized according to the key policy questions listed and discussed above. Each of the main sections of the report corresponds to one of those questions; a concluding section takes an overarching view. Each of the main sections begins with a summary of analytic approaches and findings specific to the key policy question under consideration, and then goes into depth on each of a number of subsidiary questions.

This report acknowledges extensive documentation of methods within the first annual report. A summary of these methods and new updates to them are presented in appendixes so that this report can focus on analyses completed, under way, and planned. Data sources used to conduct this multimethod evaluation include:

- American Community Survey data from the Bureau of the Census, for characteristics at the level of the census tract

- CMS Medicare and Medicaid claims and enrollment data on each beneficiary, and CMS payment data for the amount paid by CMS to each FQHC demonstration participant
- HRSA, whose Uniform Data System contains data on Section 330 grantees, including clinical measures, patient demographics, and user visits
- FQHC-level self-reported RAS data, including questions assessing progress toward becoming a PCMH and toward site-level NCQA PCMH recognition status
- TA participation reports describing FQHC-level exposure to and participation in training and other opportunities to support NCQA applications and transformation
- clinician and staff experience (CASE) surveys fielded by RAND
- baseline surveys of Medicare beneficiaries who are attributed to FQHCs or comparison sites, conducted by RAND
- interviews and focus groups involving representatives of FQHCs and PCAs involved, together with their regional organizations; these were also conducted by RAND.

Appendix A describes these data sources. Appendix B describes claims data methodologies overall, including attribution methods that assign Medicare beneficiaries to a primary care practice setting based upon the plurality of their primary care services during the 12 months prior to the demonstration. It also presents updated descriptions of our demonstration and comparison study cohorts. Appendix C documents the current status of the complete baseline beneficiary survey and the follow-up beneficiary survey currently in the field. Appendix D provides a framework for understanding the RAS and NCQA data that we use to assess site-level progress toward achieving the goal of NCQA Level 3 recognition as well as the biannual reports. Appendix E introduces methods associated with the development and fielding of the baseline CASE survey, as well as a description of the follow-up CASE survey currently in the field. Appendix F describes the methods used for our qualitative analyses.

II. Key Policy Question 1: What Are the Effects of Interventions Designed to Stimulate APC Principles on Changes in Practice Structure and on NCQA Recognition Status?

Exhibit II.1: RQs and their Derivative Subquestions for Key Policy Question 1

Section	Research Question (RQ)
II.RQ1.A	**To what extent do demonstration FQHCs become medical homes?**
RQ1.A1	How do FQHCs progress toward achievement of Level 3 recognition?
RQ1.A2	How do demonstration FQHCs achieve advances in RAS levels?
II.RQ1.B	**To what extent do demonstration FQHCs make use of the demonstration's intervention components and other supports for transformation?**
RQ1.B1	How variable are per beneficiary per quarter (PBPQ) payments and how are payments used?
RQ1.B2	How variable is the FQHC participation in NCQA TA?
RQ1.B3	How variable is the uptake of AIR TA?
RQ1.B4	How variable is the uptake of feedback reports?
RQ1.B5	How variable is site level exposure to other interventions that support APCP practices and goals?
II.RQ1.C	**What predicts FQHC use of TA and does TA uptake predict achievement of Level 3/RAS final status?**
RQ1.C1	What factors predict FQHC use of TA?
RQ1.C2	What is the association between TA uptake and achievement of Level 3 recognition/RAS final status?
RQI.C3	How important are PCA efforts?
II.RQ1.D	**How does TA support the transformation process?**
RQ1.D1	What is the quality of the TA?
RQ1.D2	How else do practices change?
RQ1.D3	What are the challenges to practice change?
RQ1.D4	Why do some FQHCs not achieve Level 3 recognition?
RQ1.D5	How else does TA help the transformation process?
RQ1.D6	How do the interventions help sites overcome challenges?
RQ1.D7	Which intervention features are most useful and which intervention features are not as helpful?

II.1 Summary of Approach to and Findings for Key Policy Question 1

CMS's APCP demonstration defines a practice's achievement of NCQA Level 3 recognition as a highly valued outcome of the demonstration. It is hoped that Level 3 recognition meaningfully signals a practice's commitment to PCMH principles and that this distinction will be associated with improved outcomes for beneficiaries associated with the practice. However, the processes by which practices successfully achieve recognition and the specific components of TA that facilitate a practice, as opposed to burdening it, are not yet known.

To address this issue, the first key policy question focuses on FQHC achievement of Level 3 NCQA status. Our analyses address demonstration site progress in achieving NCQA Level 3 recognition and in advancing RAS scores. As noted in Section I, the demonstration involves exposing participating sites to four intervention components: site PBPQ payment, TA from NCQA, TA from AIR, and access to feedback reports. We explore how demonstration sites utilize intervention components, factors that predict FQHC use of the demonstration interventions including TA, and the process of transformation.

CMS set the goal of 90 percent of demonstration sites achieving NCQA Level 3 recognition by the end of the three-year demonstration. We hypothesized that access to care management payments and TA would motivate participating sites to engage in available training sessions, webinars, conference calls, site visits, and tutorials. Furthermore we hypothesized that uptake would be associated with Level 3 recognition and transformation.

II.RQ1.A To What Extent Do Demonstration FQHCs Become Medical Homes?

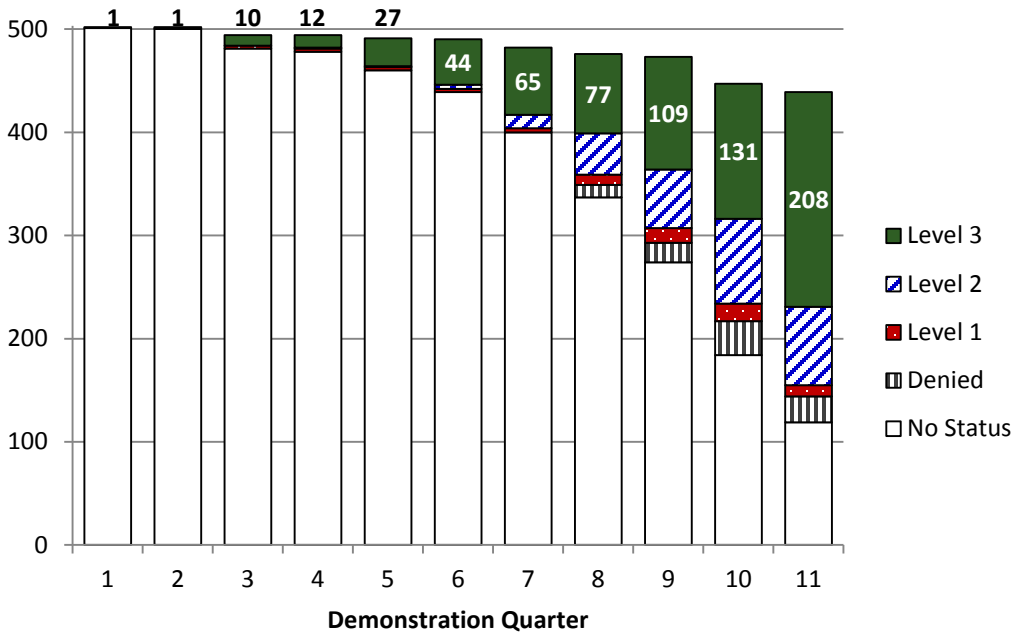
RQ1.A1 How Do FQHCs Progress Toward Achievement of Level 3 Recognition?

In November 2011, at the start of the demonstration, 500 FQHCs enrolled as participating sites, with each agreeing to the goal of achieving NCQA Level 3 recognition according to 2011 standards by the end of the planned three-year demonstration. At baseline, 29 demonstration FQHCs (5.8 percent) and 22 comparison FQHCs (2.7 percent) had 2008 NCQA Level 3; none had achieved NCQA recognition using the NCQA 2011 standards.

By August 20, 2014, three months short of the end of the demonstration, 208 of the participating 439 FQHCs (47 percent) had achieved Level 3 PCMH recognition,

76 (17 percent) had achieved Level 2 recognition, and 11 (3 percent) had achieved Level 1 recognition (Exhibit II.2). By that date, 144 participating FQHCs (33 percent) still had not achieved any level of NCQA recognition including 25 (6 percent) that were denied after application and 119 (27 percent) that either never applied or were still awaiting results.

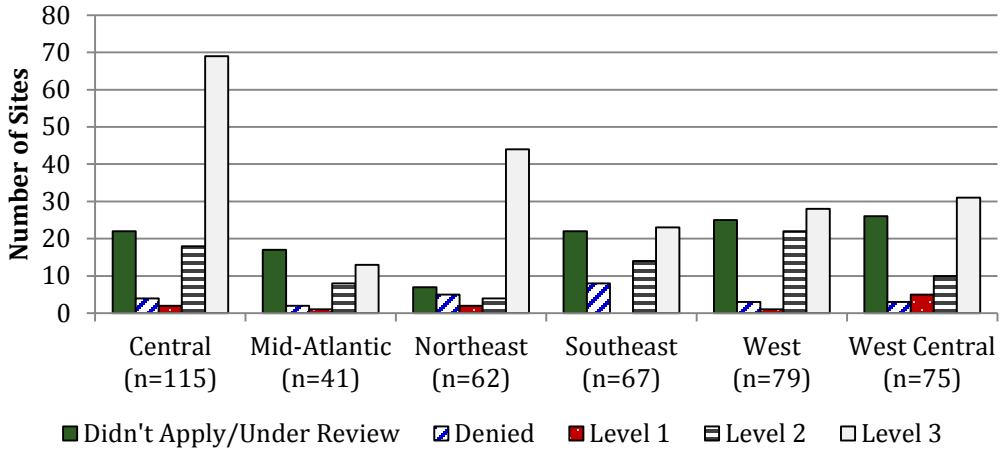
Exhibit II.2: Trends in NCQA Recognition Levels Achieved by Participating FQHCs, Demonstration Quarters 1–11 (N=439)



SOURCE: RAND analysis of CMS NCQA PCMH Recognition Status Data (current as of August 20, 2014).
NOTE: Sites labeled "No Status" include sites that have not applied for NCQA recognition and those that have applied but have yet to receive a score.

Since the demonstration was organized with multiple aspects of the TA provided by six regional PCA clusters, there is interest in examining achievement of NCQA recognition by clusters. Here we summarize NCQA recognition level by regional cluster and subsequently examine variations in regions with respect to TA uptake by FQHC sites. Exhibit II.3 shows that the Central region, comprising 115 sites, had the largest number of sites with Level 3 recognition (60 percent), while the Northeast region had the highest percentage of sites with Level 3 recognition (71 percent). Level 3 percentages were lowest for the mid-Atlantic (29 percent) and Southeast regions (33 percent).

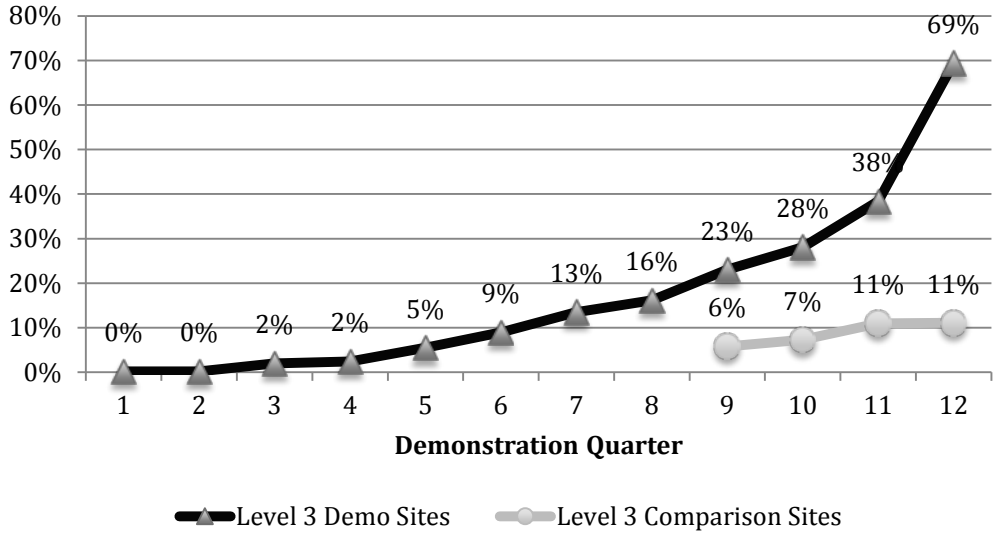
Exhibit II.3: Frequency of Site-Level NCQA Recognition Levels by Cluster (N=439)



SOURCE: RAND analysis of CMS NCQA PCMH Recognition Status Data (current as of August 1, 2014).
NOTE: Sites that have not applied for NCQA recognition or that have applied but have yet to receive a score are labeled "Didn't apply/Under review." Sites that have applied but were denied recognition are classified as "Denied."

While the bulk of this report shows data through August 20, 2014, we have been able to supplement this information with two additional reports from HRSA that include new data (through October 2014) showing important progress. As the final quarter of the demonstration was approaching, 69 percent of participating demonstration FQHCs had achieved Level 3 recognition, compared with 11 percent of comparison FQHCs. Exhibit II.4 shows trends in NCQA Level 3 recognition for demonstration and comparison sites with the sample size shown for demonstration sites restricted to the number of demonstration participants at each time point, and the sample size shown for comparison sites including the evaluation's 827 comparison FQHCs.

Exhibit II.4: Trends in NCQA Level 3 Recognition for Demonstration and Comparison FQHCs



SOURCES: Truven for demonstration sites and HRSA for comparison sites approaching the end of the demonstration's 12th quarter, October 2014. The denominator is defined for each quarter as currently participating FQHC demonstration sites.

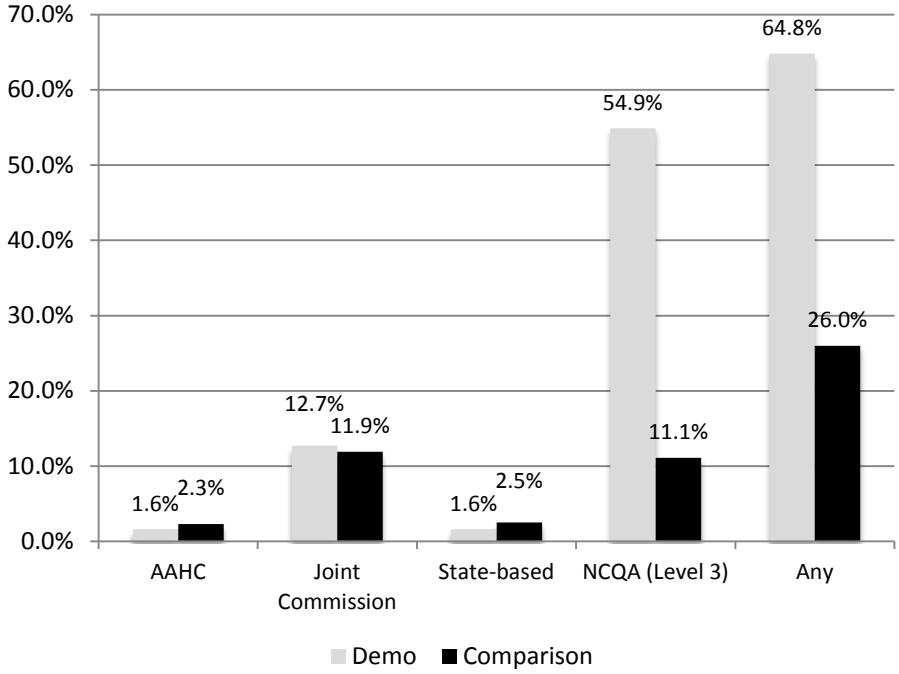
Concomitant with the demonstration, several types of PCMH recognition have been available as a means for FQHCs and other clinics to demonstrate their progress toward achieving APCP attributes. With its history of use of NCQA PCMH recognition awards since 2008, the NCQA 2011 recognition awards were the most widely distributed awards during the demonstration period. This is consistent with participating FQHCs having known since enrollment in the demonstration that they were expected to obtain Level 3 PCMH recognition from the NCQA by the time the demonstration ends.

Participating FQHCs met demonstration goals if they achieved NCQA Level 3 recognition. Additionally, all FQHC grantee organizations became eligible for supplemental funding under a HRSA program that began in 2014—independent of participation in the CMS demonstration—if the organization achieved any level of PCMH recognition from NCQA, Joint Commission or Accreditation Association for Ambulatory Health Care (AAAHC), or from the Minnesota or Oregon state-based recognition programs.[12] Exhibit II.5 compares rates of several types of PCMH recognition for demonstration and comparison FQHCs. Almost two-thirds of demonstration FQHCs (64.8 percent) compared with one-quarter of comparison FQHCs (26.0 percent) achieved any of the listed PCMH recognition types. NCQA Level 3

[12] HRSA, "Patient Centered Medical Home Supplemental Funding" website, undated. Accessed January 20, 2015: http://bphc.hrsa.gov/grants/pcmh.html

recognition was achieved by a far greater percentage of demonstration than comparison sites (54.9 percent vs. 11.1 percent, $p<0.05$), while rates were closer for AAAHC, Joint Commission, and state-based rates.

Exhibit II.5: Rates of PCMH Recognition for Demonstration and Comparison FQHCs by Source of Recognition



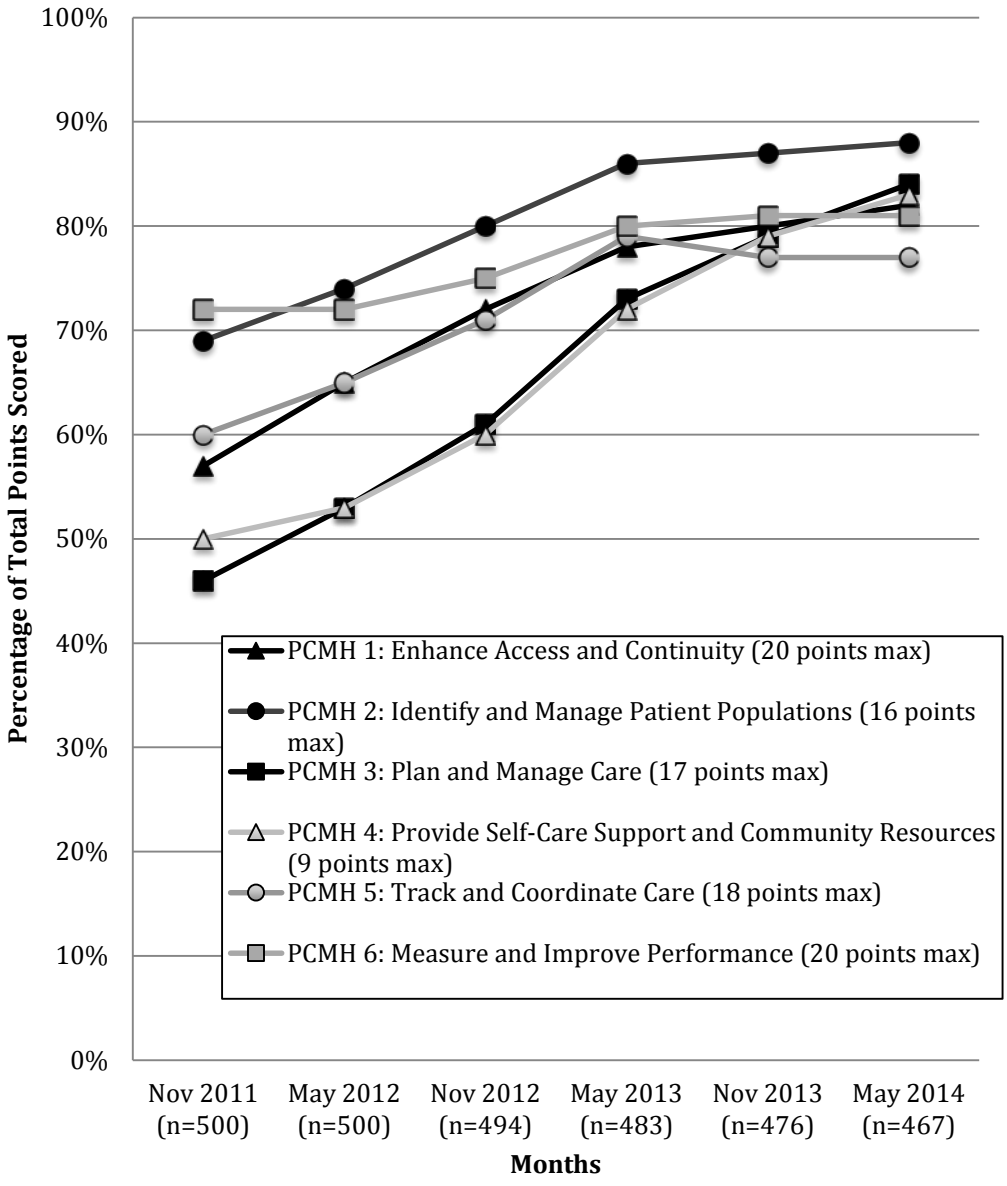
SOURCE: HRSA, October 2014.The denominator is defined as 503 demonstration and 827 comparison FQHC sites, consistent with the number of FQHC sites that respectively define the evaluations' demonstration and comparison FQHC groups.

RQ1.A2 How Do Demonstration FQHCs Achieve Advances in RAS Levels?

To understand how demonstration FQHCs achieve advances in RAS levels, we analyzed both the pattern of achievement of NCQA recognition, and also the pattern of advancement of biannual RAS scores. NCQA recognition is judged according to six domains, but detailed domain-level NCQA scores are not available to the evaluation team for either demonstration or comparison sites. In contrast, domain-specific RASs are available for demonstration sites biannually because each participating site is obligated to complete an RAS twice yearly during the demonstration. Exhibit II.6 shows the readiness of demonstration FQHCs to achieve NCQA recognition, as measured by their self-reported RAS data. Readiness has increased over time, though the rate of increase slowed

starting in November 2013 for all standards except 3 and 4. In the final 30-month assessment (May 2014), sites scored highest on their readiness to "identify and manage patient populations" (standard 2) and lowest on their readiness to "track and coordinate care" (standard 5).

Exhibit II.6: Trends in Readiness Self-Assessment Scores for Participating FQHCs, by NCQA PCMH Standard



SOURCE: RAND analysis of Truven NCQA Readiness Assessment Survey Data.

Consistent with advances in NCQA scores, RAS scores also advanced substantially from the first to the final RAS scores. Exhibit II.7 displays a cross tabulation of baseline RAS scores and current 30-month (May 2014) RAS scores. Scores are categorized according to four ranges: 0–35 (0); 36–59 (1); 60–84 (2); and 85–100 (3). At baseline, 44 demonstration FQHC sites (10 percent) had a RAS categorical score of 0, while 54 FQHC sites (12 percent) had a RAS categorical score of 3. Of those 44 sites with a baseline RAS categorical score of 0, 17 (39 percent) obtained a RAS categorical score of 3, while the remaining 27 (61 percent) obtained a lower RAS categorical score of 1 or 2. The most recent RAS scores (May 2014) showed 263 sites (60 percent of participating sites) have obtained a RAS categorical score of 3, and no sites reported a score of 0.

Exhibit II.7 also shows that not all sites starting with RAS categorical scores of 3 maintained that high level. Among the 54 sites with a baseline score of 3, only 41 (76 percent) maintained that score through the demonstration. Declining RAS scores may have resulted from errors in self-reported data, or they may reflect sites purposely correcting inflated scores in response to feedback following receipt of RAS audits. These audits are conducted on 10 percent of sites. Positively, the number of sites with low baseline RAS scores of either 0 or 1 dropped from 214 (49 percent) to 22 (5 percent) with final RAS scores.

Exhibit II.7: Frequency and Percentage of Baseline RAS Categorical Scores and Current RAS Categorical Score

	Current RAS Categorical Score (N=436*) N (Row %)				
Baseline RAS Categorical Score	**0 (0–35)**	**1 (36–59)**	**2 (60–84)**	**3 (85–100)**	**Total**
0 (0–35 score)	0 (0)	1 (2)	26 (59)	17 (39)	44
1 (36-59 score)	0 (0)	16 (9)	65 (38)	89 (52)	170
2 (60-84 score)	0 (0)	5 (3)	47 (28)	116 (69)	168
3 (85-100 score)	0 (0)	0 (0)	13 (24)	41 (76)	54
Total	0 (0)	22 (5)	151 (35)	263 (60)	436

* Three demonstration sites reported missing baseline RAS scores.
NOTE: Baseline scores were collected in November 2011 and final RAS scores were collected in May 2014.

II.RQ1.B To What Extent Do Demonstration FQHCs Make Use of the Demonstration's Intervention Components and Other Supports for Transformation?

Each of the four components of the demonstration described in Section I had the opportunity to affect the demonstration sites. This section describes how demonstration PBPQ payments, participation in TA from NCQA or AIR, and accession of site-specific feedback reports vary across FQHCs. We anticipated that sites would regularly make use of available demonstration resources to support their efforts toward APCP transformation.

RQ1.B1 How Variable Are PBPQ Payments, and How Are Payments Used?

This analysis examines PBPQ payments overall and by region, and summarizes site leader interview comments regarding how site payments are used. The median amount paid to demonstration FQHCs was $6,500 per quarter with total all-site and regional payments varying as a function of the number of Medicare beneficiaries attributed to each participating FQHC site (Exhibit II.8). The total amount paid per quarter remained consistent over time, at approximately $3.8 million.

Exhibit II.8: Medicare PBPQ Care Management Payments to Participating FQHCs

Demonstration Quarter	Total Amount Paid ($)	Median Payment per FQHC ($)	Minimum Payment per FQHC ($)	Maximum Payment per FQHC ($)
1	3,683,646	6,318	756	52,686
2	3,915,738	6,462	216	71,604
3	3,838,896	6,417	1,692	55,332
4	3,867,174	6,624	1,566	56,430
5	3,863,070	6,534	1,116	55,836
6	3,880,134	6,570	684	53,766
7	3,834,828	6,552	630	56,790
8	3,861,702	6,678	810	70,290
9	3,754,242	6,624	18	83,304
10	3,376,080	5,778	18	74,736
11	3,851,370	6,669	90	131,598

SOURCE: CMS payment data.
NOTE: The strong outlier of $131,598 in maximum payment per FQHC in the most recent quarter, which is roughly double the normal maximum, is attributable to one site and is being investigated by RAND with CMS.

Exhibit II.9 shows slightly higher total and average cumulative site payments for sites that have achieved Level 3 status compared with other sites.[13] We will continue to investigate the relationship between total (and average) cumulative site payments by NCQA level once the demonstration's final NCQA recognition levels have been determined (which is expected January 2015).

Exhibit II.9: Site Payments by NCQA Recognition Status (Level 3 vs. Not Level 3)

NCQA Level	N	Total Cumulative Site Payments by NCQA Level, Over 11 Quarters	Average Cumulative Site Payments by NCQA Level, Over 11 Quarters
Not Level 3	231	$18,682,578.00	$80,876.96
Level 3	208	$19,294,470.00	$92,761.88

Site leader interviews confirmed that sites appreciated the CMS payments; however, they often emphasized that transformation occurs at the site level, not for individual participant groups. With Medicare beneficiaries averaging 8 percent of site-level patients, site leaders described the demonstration payments as small compared with the costs needed to transform their clinics for all (not only Medicare) patients.

Across the 20 site leader interviews, analyses did not reveal a single consistent pattern in how the payments were used by sites. Leaders of multisite grantee units frequently noted that funds they received were shared across all of the grantee's clinics to facilitate transformation efforts. Appendix G (how site payments are used) provides additional thematic analysis findings about how site payments are used.

RQ1.B2 How Variable Is the FQHC Participation in NCQA TA?

TA delivered by NCQA was envisioned as a key component of the demonstration for enabling FQHCs to achieve NCQA Level 3 recognition. NCQA was expected to be responsible for enhancing demonstration site access to and uptake of webinars and other tools for supporting FQHC transformation to a PCMH.

NCQA had a slow start in effectively engaging sites in TA. As a corollary, site-level participation in TA activities (such as webinars) occurred later and at lower rates than expected. Additionally, consistent tracking of the dissemination of TA to sites and uptake of TA by sites was not systematized until midway into the demonstration. This limited the evaluation team's efforts to measure participation in TA by sites.

[13] *Cumulative payments* refers to all site payments to participating FQHCs throughout the demonstration, from the demonstration initiation and ultimately through its conclusion. The analysis described compares cumulative payments for sites that do achieve NCQA Level 3 recognition with cumulative payments for sites that do not achieve NCQA Level 3 recognition.

Participation in NCQA Webinars

At the start of the demonstration, FQHC participation in three NCQA webinars was described by NCQA and AIR as important for successful achievement of Level 3 recognition.[14] Despite minimal participation by sites early in the demonstration, site participation began to increase by the middle of the second year but remained variable throughout the demonstration. By the 11th quarter of the 12-quarter demonstration, between 68 percent and 90 percent of demonstration FQHC sites across six regional FQHC clusters did not participate in any of these NCQA webinars. No region achieved webinar participation of more than 40 percent for any of NCQA's three training webinars. (See Appendix H: NCQA Webinar Participation by Cluster).

RQ1.B3 How Variable Is the Uptake of AIR TA?

AIR served as CMS's subcontractor for structuring the delivery of TA to demonstration sites, in addition to serving as the organization responsible for tracking site use of TA, such as AIR Office Hour webinars. Participation in these webinars was limited, with less than 20 percent average participation in the 37 office-hour webinars that were delivered from September 2012 through July 2014.[15] As the demonstration approached its final quarters, we anticipated sites would have additional motivation to participate in office-hour webinars that assisted sites in the completion of NCQA applications.

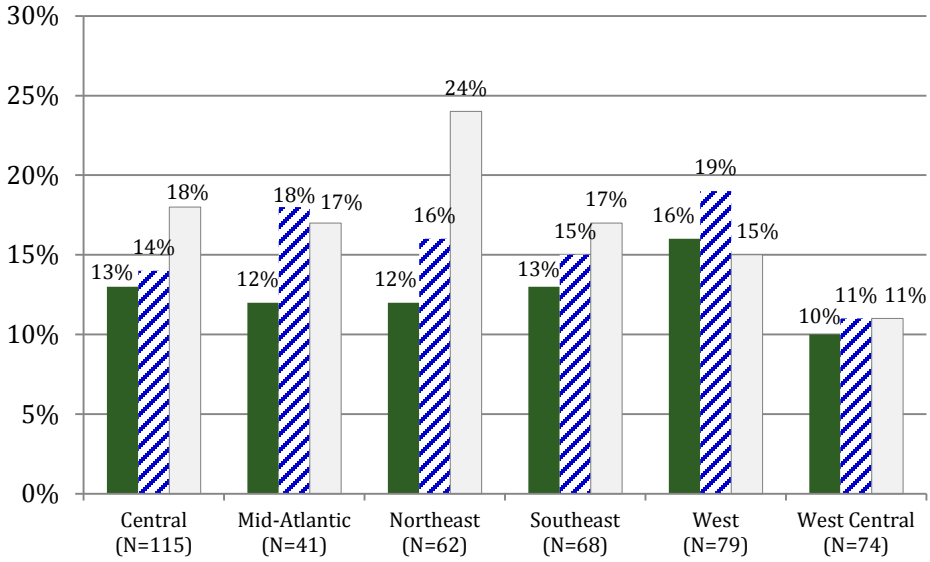
Exhibit II.10 displays average participation in all 37 AIR office-hours webinars (dark bars) compared with average participation in the 13 webinars during the most recent six-month period for which we have data prior to this report (February 2014 to July 2014). For all clusters, participation increased during the most recent six months (middle tone bars). The increase is most notable for sites that have not yet achieved NCQA Level 3 (lightest tone bars). The increase during the last six months was highest in the Northeast region, especially among the sites that had not yet achieved Level 3 recognition.

---

[14] These three NCQA webinars were PCMH Standards Part 1 and 2 and the Interactive Survey System Training.

[15] Data presented pertain to site-level participation in live office-hour webinars, though sites may have also participated in downloaded office-hour webinars using the FQHC portal. Data were not available for site-level participation in downloaded office hours except for quarters 4–7, when AIR also tracked site webinar viewing via the FQHC portal (not shown). For most clusters, the rates of download viewing were low and did not increase the total participation rates. However, adding downloaded viewing to live viewing increased participation for the Central cluster in Quarter 4 (from 12 percent to 19 percent) and in Quarter 6 (from 58 percent to 63 percent). For the West cluster, compared to live viewing only, combined live plus downloaded viewing increased in Quarter 4 (from 4 percent to 6 percent) and in Quarter 6 (from 36 percent to 38 percent).

Exhibit II.10: Average Participation in AIR Office-Hour Webinars by Participating FQHCs Stratified by Regional Cluster



SOURCES: AIR TA participation reports provided to RAND by AIR, August 8, 2014; Truven Recognition data provided to RAND on December 18, 2013, for starting the six-month window.
NOTE: Darkest bars represent average participation in all office-hour webinars; middle tone bars represent average participation during the last six months, and lightest tone bars represent average participation during the last six months only among sites that had not yet achieved NCQA Level 3 recognition.

RQ1.B4 How Variable Is the Uptake of Feedback Reports?

With its subcontractors, CMS made available three types of site-specific reports to participating sites through the FQHC data portal. These reports provided clinic- or beneficiary-level data with potential to enhance sites' NCQA recognition application or their efforts toward transformation.[16] It has been hypothesized that site-level review of these feedback reports could serve as a marker for site-level exposure to and use of the

[16] Beginning in Quarter 6, FQHCs received three types of reports via the FQHC data portal to motivate transformation toward becoming a medical home. First, the biannual NCQA RAS Report provides FQHCs with current site-level NCQA PCMH recognition level and overall score trends. Second, the quarterly cost and utilization data reports provide site-level claims-based utilization measures (e.g., inpatient admission, ED visits), Medicare expenditure summary data (e.g., average total Medicare expenditure per beneficiary), and quality of care measures (e.g., glycated hemoglobin blood [HbA1c] testing, retinal eye exams, low-density lipoprotein [LDL] screening and nephropathy testing rates among beneficiaries with diabetes). Third, the quarterly claims-based beneficiary-level report provides a file summarizing key study outcomes for all beneficiaries attributed to the FQHC (e.g., cost, utilization, health data).

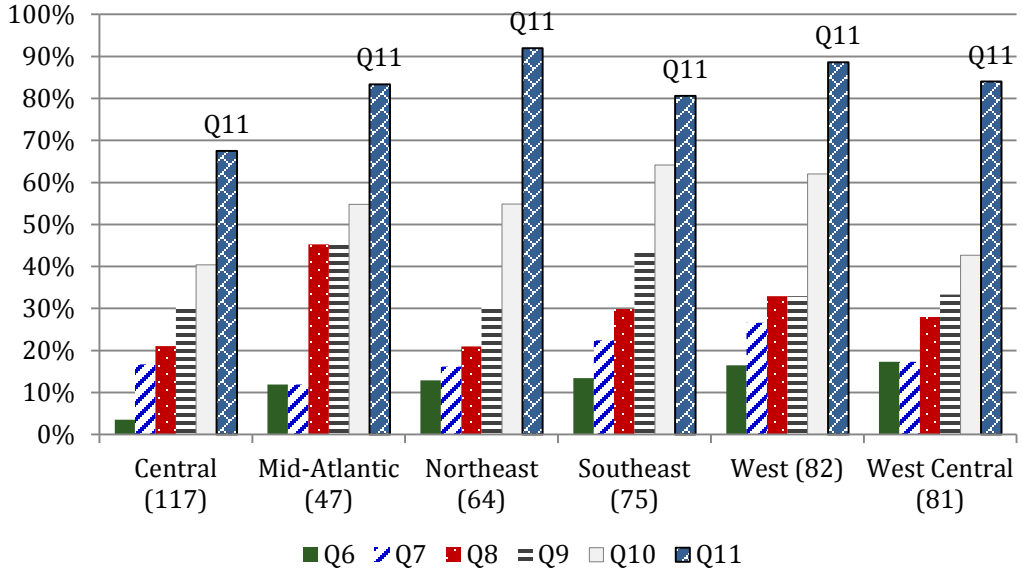
demonstration's interventions, since the dates of download are recorded for all participating sites.

Feedback reports became available during demonstration Quarter 6, but sites only logged on infrequently to see any of these three reports. Exhibit II.11 shows that even by Quarter 10, only slightly more than half (52 percent, n=227) of participating sites within each of the six regional clusters had *ever* logged into the portal (even once) to view any of their feedback report data.

When RAND discussed low rates of feedback reports with sites and with PCA leaders, we learned that many sites did not understand the potential value these reports could have for facilitating either their NCQA application or their transformation. Through discussions with AIR, PCA leads, and CMS, methods for accessing these reports and for using them to support site transformation efforts were more widely disseminated.

By the end of Quarter 11, 81 percent of sites (with a range of 68 percent to 92 percent of participating sites within each regional cluster) had logged into the portal to view at least one of their three feedback reports. Between 8 percent and 22 percent of sites within regions never downloaded even one feedback report.

Exhibit II.11: Utilization of Feedback Reports Over Time, Showing Percentage of Region with Any Portal Log-In to View a Report (N=439)



RQ1.B5 How Variable Is Site Level Exposure to Other Interventions That Support APCP Practices and Goals?

A provision of participating in CMS's APCP demonstration was that sites were able to join additional initiatives and demonstrations focused on redesigning care delivery or providing supplemental funding to support PCMH transformation. While not truly an intervention of the demonstration, it is notable that HRSA provides additional financial and infrastructure support to participating FQHCs to cover the cost of applying for NCQA PCMH recognition and also provides some start-up funds to cover the costs associated with transforming into an APCP.

To assess the prevalence of sites taking advantage of these opportunities, we count site-level participation in each of three domains: CMS-funded activities, HRSA-funded PCMH-related activities, and other, externally funded PCMH-related activities.

Many demonstration and comparison FQHCs participate in initiatives pertinent to PCMH transformation. Exhibit II.12 shows that nearly 28 percent of demonstration FQHCs participate in a CMS PCMH transformation demonstration beyond the APCP demonstration, while only 22 percent of comparison FQHCs participate in a CMS transformation initiative. Ninety-six percent of demonstration and 91 percent of comparison site FQHCs participate in at least one HRSA transformation initiative; most participate in several. Additionally, a small number of demonstration and comparison sites participate in other relevant initiatives.

Exhibit II.12: Counts of FQHC Participation in Other CMS, HRSA, or Other PCMH or Quality Improvement Initiatives Pertinent to Transformation

	Demonstration Sites		Comparison Sites	
Rate of Response	N	Percentage	N	Percentage
Participation in CMS demonstrations[a]				
None	364	72.4	645	78.0
Any	139	27.6	182	22.0
Participation in HRSA initiatives[b]				
None	20	4.0	74	8.9
1	84	16.7	191	23.1
2	147	29.2	226	27.3
3	136	27.0	230	27.8
4	85	16.9	83	10.0
5	15	3.0	20	2.4
6	16	3.2	3	0.4
Participation in other initiatives[c]				
None	491	97.6	818	98.9
Any	12	2.4	9	1.1

CMS demonstrations indicate site-level participation in any of three CMS initiatives: Pioneer, Medicare

Shared Savings Plan, and the North Carolina 646 Demonstration.
[b] HRSA initiatives indicate site-level participation in any of the following programs: Beacon Communities Supplemental Funding; Affordable Care Act (ACA) grants, including: Building Capacity, ACA Immediate Facility Improvement, and New Access Point; ARRA grants; or the HRSA PCMH Initiative. Data are current as of June 2013.
[c] Other initiatives indicate site-level participation in the Safety Net Medical Home Initiative.

II.RQ1.C What Predicts FQHC Use of TA and Does TA Uptake Predict Achievement of Level 3/ RAS Final Status?

To better understand reasons for variable uptake of TA, we assessed factors associated with its use and analyzed the impact of its uptake by assessing the association between uptake and achievement of Level 3 recognition/RAS final status.

We conducted separate logistic regression analyses to estimate the adjusted association of site- and area-level characteristics (primary predictors) on the likelihood of TA uptake (dependent variables) and the adjusted association between TA use (primary predictors) and Level 3 recognition (dependent variable). We repeated this for four different types of TA:

- attending five or more AIR office hours (among demonstration sites that attended at least one office hour)
- viewing five or more RAS, cost, and utilization feedback reports (among demonstration sites that viewed at least one report)
- attending five or more AIR webinars (among demonstration sites that attended at least one webinar)
- attending at least one NCQA webinar (among all demonstration sites)

For both types of multivariable analyses, those that examined factors associated with TA participation and those that used AIR office hours, webinar attendance, and feedback report viewing as the primary predictor variables, the sample of sites was limited to those that had used that particular type of TA at least once. All analyses excluded sites that had achieved NCQA Level 3 recognition prior to the start of TA. Among sites that achieved NCQA Level 3 recognition after the start of TA, counts of TA attendance were limited to those observed prior to NCQA recognition. These results are preliminary and will be repeated when the final set of Level 3 recognition outcomes are transferred to RAND from Truven.

RQ1.C1 What Factors Predict FQHC Use of TA?

We examined the impact of site- and area-level characteristics on the likelihood that demonstration sites use TA. Results are shown as odds ratios (ORs), where an OR greater than 1 with a p-value of less than 0.05 for a particular characteristic indicates that the characteristic independently increases the likelihood that a site participates in a particular type of TA at least five times (for attending AIR office hours and webinars and viewing feedback reports).

Exhibit II.13 shows that a site's baseline RAS score is an important factor predicting participation in TA. Across all measured types of TA, sites with lower baseline RAS scores are more likely to participate in TA compared with sites that have higher baseline RAS scores. For example, among sites that participated in at least one AIR webinar, sites with RAS Level 0 at baseline had statistically significant increased odds (OR=1.55) of attending five or more webinars compared to sites with RAS Level 1 at baseline, while sites with RAS Level 2 at baseline were significantly less likely to participate (OR=0.43). Similarly, among sites that viewed at least one feedback report, sites with RAS Level 0 at baseline had statistically significant increased odds (OR=2.91) of viewing five or more feedback reports compared to sites with RAS Level 1 at baseline; sites with RAS Level 3 at baseline were significantly less likely to view five or more feedback reports (OR=0.21).

Site characteristics that indicate a higher level of resources at baseline (receiving PCMH payments from one or more APCP-related plans or total site revenue)[17] were associated with an increased likelihood of participating in AIR office hours or webinars. Specifically, among sites that attended at least one AIR webinar, those with higher revenues were more likely to attend five or more webinars (OR=1.66 for each additional million dollars in revenue). Among sites that attended at least one office hour, those receiving PCMH incentive payments from one or more plans (beyond the APCP demonstration) were more than twice as likely to attend five or more office hours (OR=2.5).

[17] The demonstration's APCP application asks a question: "Are you currently receiving payments from payers *other than Medicare* to be a medical home for any of your patients?" FQHC sites were then asked to list in free text boxes which payers provided these payments and the percentage of patients to which these payments applied. We used this variable to assess baseline site receipt of PCMH payments from one or more APCP-related plans. Total site revenue data are grantee-level data from the Uniform Data System and are measured as the total revenue divided by the number of delivery sites.

Exhibit II.13: Multivariate Relationship Between Site- and Area-Level Characteristics and TA Uptake

	Measure of TA Participation					
	Attending 5+ AIR Office Hours (among those who attended at least one) (n= 384)		Viewing 5+ Feedback Reports (among those who viewed at least one) (n= 315)		Attending 5+ AIR Webinars (any type, among sites who attended at least one) (n=208)	
Characteristic	OR	Standard Error (SE)	OR	SE	OR	SE
Baseline RAS Level 0 (<35 points)	1.24	0.29	2.91**	0.46	1.55**	0.28
Baseline RAS Level 1 (35–59 points)	[ref]					
Baseline RAS Level 2 (60–84 points)	0.55	0.22	0.81	0.32	0.43**	0.19
Baseline RAS Level 3 (85–100 points)	0.65	0.33	0.21**	0.49	0.47	0.30
Ambulatory Care Quality Accreditation	0.33***	0.33	0.68	0.45	1.00	0.25
Receiving PCMH payments from one or more plans	2.50*	0.39	0.86	0.47	0.89	0.31
Total revenue per site (in millions)	1.04	0.23	0.73	0.32	1.66*	0.22
Participation in CMS Shared Savings Demo	0.78	0.34	1.43	0.44	0.52*	0.30
Number of service delivery sites: 1–10 sites	[ref]					
Number of service delivery sites: 11+ sites	1.44	0.14	0.38*	0.19	0.91	0.13
Rural-Urban Continuum Code: Metro	0.81	0.19	0.89	0.27	0.87	0.17
Nonmetro–rural	0.77	0.26	1.35	0.38	0.87	0.23
Nonmetro–urban	[ref]					
PCA Region: Central	0.77	0.25	0.13*	0.37	0.37	0.21
Mid-Atlantic	0.74	0.37	0.07**	0.54	0.34	0.35
Northeast	0.67	0.33	0.66*	0.38	1.00*	0.30
Southeast	0.58	0.31	0.21	0.43	0.29	0.28
West	0.53	0.29	0.49	0.32	0.39	0.25
West-Central	[ref]					

NOTE: * p<0.05; ** p<0.01; *** p<0.001;

These analyses controlled for a number of additional site- and area-level characteristics, including the number of service delivery sites operated by the site's grantee organization, whether the sites are located in urban or rural areas, and an indicator of the PCA region that oversees the delivery of some forms of TA to the site. Additional characteristics were explored in preliminary analyses but were removed from the final model due to a lack of association with TA participation. These included: mean site-level measures of beneficiary characteristics, percentage of patients at the site that have Medicaid or are uninsured, the number of years the FQHC has been in operation, the number of Medicare beneficiaries attributed to the site, counts of primary care and specialist providers, and the percentage of household poverty in the census tract in which the FQHC operates, as well as indicators of whether the site is participating in the HRSA PCMH Initiative, is a Health Center Controlled Network (HCCN) grantee, and the Medicaid program in the state differentially pays FQHCs based on PCMH recognition status.

Many of the site-level factors that are likely to be important predictors of TA use are not known consistently across demonstration sites, and, therefore, were not available for inclusion in these analyses at the site level. These factors include the sophistication of the site's electronic health record (EHR) system, the numbers of nonclinician staff available to participate in TA, the strength of the FQHC's practice leadership, its leaders' priorities, the extent to which information about the demonstration is communicated from management to staff, the level of encouragement and time provided by leadership to participate in TA, and the site's practice culture (including how strongly staff embrace quality improvement initiatives). Many of these factors have been examined through RAND's qualitative analyses. Additionally, we are working on evaluating whether baseline CASE reports by clinicians and staff predict TA uptake.

A similar pattern of results was noted between site- and area-level characteristics and participation in at least one NCQA webinar. (See Appendix I: Predictors of NCQA Webinar Participation.)

RQ1.C2 What Is the Association Between TA Uptake and the Achievement of Level 3 Recognition/ RAS Final Status?

We also examined the extent to which TA participation is associated with Level 3 PCMH recognition, adjusting for site- and area-level characteristics. These analyses use logistic regression to estimate the adjusted association between TA use (primary predictor variable) and Level 3 recognition (dependent variable), which we define as a total RAS score of 85 points or higher on the 30-month RAS survey. We derived Level 3 recognition in this manner because RAS scores are likely to provide a more accurate

measure of a site's current PCMH *readiness* as compared with a site's current NCQA *recognition status*. This assumption is based on the observation that many demonstration sites have elected to delay their application for formal recognition until the end of the demonstration. Thus, many sites that lack Level 3 recognition actually have RAS total scores of 85 points or higher that might qualify the site for recognition if it were to apply.

We used the same types of TA that we used to estimate predictors of TA uptake (See Section II.RQ1.C) and limited the sample of sites noted as using TA to the subset of those that used the particular type of TA at least once. All analyses excluded sites that had achieved NCQA Level 3 recognition prior to the start of TA. Among sites that achieved NCQA Level 3 recognition after the start of TA, counts of TA attendance were limited to those observed prior to NCQA recognition. These results are preliminary and will be repeated when the final set of Level 3 recognition outcomes are transferred to RAND from Truven.

In three separate regression analyses, preliminary results show that there is no statistically significant association between

- attending five or more AIR webinars and achieving a Level 3 score on the 30-month RAS among sites that attended any AIR webinars (Exhibit II.14, column 1)
- viewing five or more feedback reports and achieving a Level 3 score among sites that viewed at least one feedback report (Exhibit II.14, column 3)
- attending at least one NCQA webinar and achieving a Level 3 score, among all sites (Exhibit II.15).

Preliminary results show that there is a negative association between AIR office-hour attendance and achieving a Level 3 RAS score. Among sites that attended at least one AIR office hour, sites that attend five or more were far less likely to achieve a Level 3 RAS score (OR=0.55) (Exhibit II.14, column 2). These results are preliminary and should be interpreted with caution until this analysis is repeated with final NCQA recognition values. As noted above, there is a strong association between baseline RAS level and TA participation; sites with lower baseline RAS levels are more likely to use TA throughout the demonstration (See Exhibit II.13). Exhibits II.14 and II.15 show that sites with lower baseline RAS scores are also less likely to achieve a Level 3 RAS score at 30 months. Although all models controlled for baseline RAS in these analyses, there is a strong possibility that residual confounding may bias these estimates. Because sites that use more TA are also more likely to have lower-level baseline RAS, they may be less likely to achieve a Level 3 score by 30 months because of their lower baseline readiness and not because of their use of TA. In fact, the interactive nature of AIR's office-hour webinars

may mean it is the TA modality most likely to be used by struggling sites. Furthermore, for sites with low RAS scores, the TA offered may not have been sufficient to support the transformation to Level 3 recognition.

Exhibit II.14: Relationship Between TA Participation (AIR Webinar and Office Hour Participation and Feedback Report Viewing) and 30-Month RAS Level 3 Status

	Measure of TA Participation					
	AIR Webinar Attendance (Low=1–4, High=5+) (n= 384)		AIR Office Hour Attendance (Low=1-4, High=5+) (n= 315)		Feedback Report Viewing (Low=1-4, High=5+) (n= 208)	
Characteristic	**OR**	**SE**	**OR**	**SE**	**OR**	**SE**
Low TA use (1-4)	[ref]					
High TA use (5+)	1.06	0.24	0.55*	0.27	1.01	0.33
Baseline RAS Level 0 (<35 points)	0.72	0.37	0.71	0.42	1.67	0.53
Baseline RAS Level 1 (35–59 points)	[ref]					
Baseline RAS Level 2 (60–84 points)	2.63***	0.26	2.13**	0.29	2.56**	0.36
Baseline RAS Level 3 (85–100 points)	2.74*	0.45	3.90**	0.49	12.01**	0.80
Ambulatory Care Accreditation	1.03	0.26	0.54*	0.31	0.86	0.42
HCCN grantee	1.16	0.23	1.12	0.26	1.06	0.33
Medicaid payments to PCMHs under way	1.42	0.25	1.27	0.27	0.96	0.33
Receiving PCMH payments from one or more payers	0.76	0.35	0.51	0.46	0.68	0.51
Medicare beneficiaries attributed in baseline year	1.00	0.00	1.00	0.00	1.00	0.00
Total revenue per site (in millions)	1.04	0.21	1.38	0.24	1.50	0.32
HRSA PCMH Initiative participant	0.69	0.24	0.61	0.26	0.90	0.32
Participation in CMS Shared Savings Demonstration	3.07***	0.29	3.06**	0.33	2.30*	0.41
Number of service delivery sites: 1–10 sites	[ref]					
Number of service delivery sites: 11+ sites	0.76	0.26	0.80	0.29	0.85	0.35
Rural-Urban Continuum Code: Metro	1.63	0.29	2.18*	0.33	1.58	0.40
Nonmetro–rural	1.55	0.39	1.92	0.45	1.12	0.58
Nonmetro–urban	[ref]					
Percentage household poverty in census tract	1.00	0.01	1.00	0.01	0.99	0.02

* p<0.05; ** p<0.01; *** p<0.001

Exhibit II.15: Relationship Between NCQA Webinar Participation and 30-Month RAS Level 3

Characteristic	OR	SE
Attending no NCQA webinars	[ref]	
Attending at least one NCQA webinar	0.69	0.22
Baseline RAS Level 0 (<35 points)	0.62**	0.27
Baseline RAS Level 1 (35–59)	[ref]	
Baseline RAS Level 2 (60–84 points)	2.37**	0.18
Baseline RAS Level 3 (85–100 points)	2.47*	0.28
Ambulatory Care Accreditation	0.94	0.24
HCCN grantee	1.08	0.22
Medicaid payments to PCMHs under way	1.72*	0.23
Receiving PCMH payments from one or more payers	0.72	0.34
Medicare beneficiaries attributed in baseline year	1.00	0.00
Total revenue per site (in millions)	1.22	0.20
HRSA PCMH Initiative participant	0.73	0.22
Participation in CMS Shared Savings Demo	2.75***	0.28
Number of service delivery sites: 1–10 sites	[ref]	
Number of service delivery sites: 11+ sites	0.84	0.12
Rural-Urban Continuum Code: Metro	1.61	0.16
Nonmetro–rural	1.66	0.22
Nonmetro–urban	[ref]	
Percent household poverty in census tract	1.00	0.01

NOTE: A total of 431 demonstration sites were included in this analysis.
* p<0.05; ** p<0.01; *** p<0.001

One strong site-level predictor of achieving a Level 3 score on the 30-month RAS was a site's participation in a CMS shared-savings demonstration.[18] Sites participating in these demonstrations had a likelihood of achieving Level 3 status by 30 months that was two to three times higher than sites not participating in these demonstrations—an association that was found in all models that examined the impact of TA on achieving Level 3 RAS scores. Sites that participate in shared-savings demonstrations may benefit from additional resources gained from participating in these demonstrations, which facilitates achievement of Level 3 recognition. In addition, the potential for shared savings may serve as a potent driving force behind these sites' pursuit of Level 3 recognition.

These analyses controlled for a number of additional site- and area-level characteristics, including whether the site receives PCMH payments from one or more

[18] Among FQHCs, 214 participated in the Medicare Shared Savings Program, 60 participated in Pioneer, 16 participated in the Medicare Health Care Quality Program, and one participated in the Multi-Payer Advanced Primary Care Demonstration.

payers, the number of beneficiaries per site, total revenue at the site level, whether the site is participating in the HRSA PCMH initiative, ambulatory care quality accreditation, the number of service delivery sites operated by the site's grantee organization, the extent to which the site is located in an urban or rural area, an indicator of whether the site is located in a state whose Medicaid program makes incentive payments based on an FQHC's PCMH status, and the percentage of household poverty in the census tract of the FQHC. None of these site- or area-level variables were consistently associated with achieving a Level 3 score on the 30-month RAS across models.

Additional covariates were explored in preliminary analyses but were removed from the model due to a lack of association. These included: mean site-level beneficiary characteristics, percentage of patients at the site who are enrolled in Medicaid or are uninsured, the number of years the FQHC has been in operation, counts of primary care and specialist providers, and indicators of the PCA region that oversees the delivery of some forms of TA to the site.

As with the multivariable TA analyses described above, these analyses are also limited by the type of site-level factors that are available consistently across sites. It is likely that other site-level factors that are not available to us—particularly those related to site resources, practice culture, and climate—are associated with progressing toward PCMH recognition. Some of these will be explored in the survey and qualitative analyses.

In addition to the caveats noted above about the lack of available site-level factors across sites, since the initiation of this evaluation, RAND has noted the lack of robust metrics for assessing TA across sites and across time. With the help of AIR and CMS, we have settled on a set of variables for defining TA consistently across time and sites. Nevertheless, we recognize that these metrics do not consistently measure the entire demonstration period. Earlier in the demonstration, AIR had several other systems for measuring TA exposure and uptake. Despite our measurement of TA delivered by PCAs and taken up by sites, we acknowledge that the available metrics may not adequately assess the TA delivered or used. Accordingly, we will benefit substantially from our multimethod analyses that supplement quantitative analyses with qualitative ones.

RQ1.C3 How Important Are PCA Efforts?

AIR contracted with regional PCAs that in turn worked with state PCAs to structure the dissemination of TA to participating FQHCs. While AIR and PCAs agreed to a set of reporting criteria for tracking state-level PCA leadership of TA activities across the demonstration sites within their states, the metrics and frequency of their collection changed multiple times during the first two years of the demonstration. Beginning in July

2013, state PCAs regularly reported five types of PCA leadership activities for each of four consecutive quarters: webinars, conference calls, site visits, phone calls, and emails.

Exhibit II.16 summarizes TA delivered to participating sites by their state PCAs. Slightly more than one-third of participating FQHCs participated in their state's PCA-led webinars across all four quarters, though one-fourth of the sites never experienced any state PCA–led TA webinars. Sites varied with respect to their experience with their state PCAs conducting site visits regarding TA: 14 percent of FQHC sites never participated in a state-led site visit while 17 percent were involved in site visits for all four quarters. Sites also varied with respect to their participation in state PCA–led conference calls regarding TA. Almost one-fourth of states never participated in state PCA–led TA conference calls while another 26 percent participated in these conference calls for at least three of the four quarters. Almost all sites (97 percent) received their state's PCA-led email pertinent to TA across all four quarters and 82 percent of sites received their state's PCA-led phone calls relevant to TA across all four quarters.

Exhibit II.16: Participating FQHC Site Report of State PCA Leadership of TA by Modality (N=439)

	Number of Quarters Reported (n=439) N (Row %)									
TA Activity	**0**		**1**		**2**		**3**		**4**	
Email	3	(<1)	0	(0)	0	(0)	11	(3)	425	(97)
Phone call	3	(<1)	0	(0)	24	(5)	50	(11)	362	(82)
Webinar	107	(24)	82	(19)	77	(18)	21	(5)	152	(35)
Site visit	60	(14)	74	(17)	99	(23)	132	(30)	74	(17)
Conference call	99	(23)	121	(28)	103	(24)	80	(18)	36	(8)

Overall, by the 11th quarter of the demonstration, the state PCAs associated with the mid-Atlantic region outperformed other regions across multiple types of activities, though the West and West Central regions also demonstrated high levels of leadership activities.

We found no association between state-reported PCA leadership activities and site-level participation in webinars or accession of feedback reports. We also found no association between state-reported PCA leadership activities and the proportion of regional cluster sites that achieved NCQA Level 3 recognition. Appendix J (PCA Delivery of TA) shows additional detail regarding the progression over time of state PCA leadership in five types of TA for which they were responsible using reporting criteria developed jointly by AIR and PCAs. Appendix J also shows variations in state leadership by regional cluster.

II.RQ1.D How Does TA Support the Transformation Process?

RQ1.D1 What Is the Quality of the TA?

One means of assessing the quality of the TA is to ask clinicians and staff associated with demonstration sites whether TA components were clear and useful. The CASE survey supported this assessment. Among the approximately 44 percent of responding physicians, nurse practitioners, and physician assistants who reported attending a webinar or training session about access, more than 90 percent reported that the presentation of information was clear, and more than 80 percent reported that the *information was useful*. These ratings did not vary by RAS score, but the percentage reporting high usefulness was statistically significantly lower in nonmetropolitan rural sites, suggesting that the content of training about access may have been somewhat less relevant to rural FQHCs. Similar findings were reported about the clarity and usefulness of webinars or training sessions about care coordination.

RQ1.D2 How Else Do Practices Change?

To better understand how practices change, we examined CASE data beyond the analysis of the clarity and usefulness of TA. The CASE survey queried respondents who attended webinars or training sessions about how (if at all) providers and practices *changed their behaviors* in relation to their TA exposure. As an example, among the 44 percent of responding physicians, nurse practitioners, and physician assistants who reported attending a webinar or training session about access, approximately two-thirds reported that these webinars or training sessions *changed the way patients schedule appointments with their practices;* slightly fewer indicated these training sessions *changed the ways patients can contact providers in the practice* (meaning increased access for patients to contact their clinicians through expanded phone and email access).

Similarly, among those who reported attending a webinar or training session about care coordination, approximately 70 percent of responding physicians, nurse practitioners, and physician assistants reported that these webinars *changed the way site providers in their practice communicate with each other.* Approximately 60 percent reported *changes in the way site providers communicate with specialists, hospitals, or EDs*. These responses did not vary by RAS scores. While the baseline CASE survey alone cannot measure practice changes over time, cross-sectional relationships between RAS scores and measures of practice climate may foreshadow changes that will occur as sites increase their RAS scores over time and attempt to receive NCQA Level 3 recognition as medical homes.

Among those who had seen a feedback report about becoming a medical home, more than 80 percent of responding physicians, nurse practitioners, and physician assistants reported that *their work had changed in response to the reports*, and almost 85 percent reported that the *work of others had changed*. These ratings did not vary by RAS score, with the exception of lower likelihood of the respondent's own work changing at RAS Level 3 than at RAS Level 1 in unadjusted analysis (losing statistical significance in adjusted analysis).

Among the responding physicians, nurse practitioners, and physician assistants who had seen feedback reports that compared their sites to other practices on measures of health care utilization, costs, and quality, approximately 70 percent of responding physicians, nurse practitioners, and physician assistants reported *changes in the work they and others at their sites performed*. These reports of changes in work did not vary by RAS score.

Among those who had seen reports of hospital and ED use by Medicare beneficiaries at their sites, more than half reported that these lists were used to contact patients after a hospitalization, to contact patients after an ED visit, and to change the way all patients (including patients who were not Medicare beneficiaries) receive care. Compared with RAS Level 1 sites, respondents at RAS Level 2 sites were less likely to say these reports were used to change the way all patients receive care, and there was a similar trend for RAS Level 3 sites that did not reach statistical significance.

RQ1.D3 What Are the Challenges to Practice Change?

CASE survey data from participating FQHCs and thematic analyses of 30 site leader interviews, along with six PCA leader interviews and six PCA focus groups, provide insights into challenges to practice change. All but two of the demonstration FQHCs in our baseline interview sample reflected on challenges with PCMH implementation related to such general-change management issues. Overall, they echoed well-developed themes about PCMH transformation, such as the notion that PCMH transformation is closely linked to general practice improvement and redesign efforts in primary care, sharing many core processes of organizational innovation and change management. According to the Agency for Healthcare Research and Quality, a critical element of the PCMH is an overall commitment to quality, reflected in a strong culture and mindset of continuous quality improvement that supports, tracks, and maintains such activities as using evidence-based medicine and clinical decision support tools to guide shared decisionmaking with patients and families, engaging in performance measurement and

improvement, measuring and responding to patient experiences and patient satisfaction, and practicing population health management.[19]

Site leaders focused on change management challenges centered on creating buy-in across the staff, attempting to integrate or embed new tasks into current processes to reduce burden, overcoming provider reluctance to changing practice behavior, and the axiom that *"just a lot of change is required to become a PCMH."*

Specialty Access

One important set of challenges facing FQHC sites participating in the demonstration as they transform into APCPs appears to include limited access to timely specialist services outside the site and high-quality mental health services in any location.

Among CASE survey respondents, approximately 20 percent of clinicians (only one in five) reported that it was *easy to obtain timely new-patient office visits with specialists outside their sites*. This did not vary by RAS level, but respondents in nonmetropolitan urban areas were statistically significantly more likely to report such easy access.

Not only was *easy* access for new-patient office visits with specialists reported for only one in five respondents, a similar pattern of limited access was reported for specialist *follow-up visits*, for specialist *procedures*, and for *mental health provider visits*. Among CASE respondents, only 30 percent reported easy access to timely follow-up specialty visits; 23 percent reported easy access to timely specialist procedures, and fewer than 20 percent reported easy access to high-quality mental health services.

The problem of limited access to specialty services was even more notable in rural sites and more notable for the special case of access to mental health specialty services.

Qualitative analyses of site leader interviews confirm this challenge. While all sites mentioned specialty access as a main problem, none identified limited specialty access as a problem for scoring well on NCQA applications. Presumably this is because the 2011 standards require that sites have systems in place to coordinate, track, and "close the loop" on referrals. Sites are not required to document easy access for patients to see specialists in a timely manner.

Inadequate specialty access at FQHCs potentially puts beneficiaries at risk for suboptimal processes and outcomes including more ED visits, hospitalizations, readmissions, morbidity, and death. Site-level metrics are expected to show these

---

[19] Taylor, E. F., Peikes, D., Genevro, J., and Meyers, D. (2013). *Creating capacity for improvement in primary care: The case for developing a quality improvement infrastructure,* Rockville, Md.: Agency for Healthcare Research and Quality, Decisionmaker Brief: Primary Care Quality Improvement No. 1, Document No. PP13-82.

outcomes with higher site-level rates of hospitalizations, readmissions, morbidity, and death after adjustment for age and baseline burden of illness.

Staff Buy-In

FQHC respondents commonly reported challenges in creating buy-in across all the staff. They noted that the PCMH transformation process involved essentially changing the culture of the entire clinic. A few respondents specifically mentioned the challenges of changing culture around teamwork, but the majority discussed how hard it is to gain buy-in to change across the whole organization. A few of the FQHCs addressed the issue of staff buy-in by "holding full team meetings and involving all staff in the process of thinking through how a change could be implemented." In one practice, they reported that the "full turnover of all practice staff except the physicians was needed to change the culture and allow for the changes necessary to become a PCMH."

Integrating and Embedding Tasks into Routine Processes

Another general change challenge discussed in depth by some demonstration respondents involved efforts to reduce burdens on staff by integrating and embedding the required tasks for the many aspects of PCMH—previsit preparation, chart review, documenting medications, care plans, patient education, etc.—into general practice operation without having physicians or other clinical staff feel like more work was being added.

> "Our providers are already busy with taking care of patients, documenting, a lot of paperwork. So if we continue to add more tasks [with PCMH], it may be very difficult for them."

One respondent described the necessity of embedding tasks and how difficult that can be, using the example of team huddles.

> "The challenge is how you really integrate the providers' assessment of the patient and their input into the process without making it that the team has got to be there at 8:00 in the morning and stay after hours. It's kind of trying to figure out creative ways that [team huddles] are working, and this is how we make it work without making it feel like it is additional work. I mean, you can only really accomplish that if you make it integrated into the work such that it just is getting done versus that it's an extra activity that we're trying to accomplish."

Provider Reluctance to Change

Another prevalent issue was providers' reluctance or resistance to change, a widespread issue in health care improvement and redesign. "Provider reluctance" was described as foot-dragging, or hesitancy to change, rather than outright opposition or

resistance. Some respondents perceived generational differences in the reluctance or dissatisfaction with taking part in the initiatives to document quality or use of EHRs to document processes of care. Others attributed reluctance more to individual providers' comfort with change.

> "The providers, most of them are committed. But since they have a lot of demands, some providers find it difficult to meet the requirements and do the steps needed."

Two strategies for addressing this challenge included taking an "incremental approach to change, using baby steps" and framing the PCMH effort as a solution to common frustrations, such as managing hospital referrals and poor coordination of care; e.g., PCMH will solve physician frustration with access to specialty care.

Many FQHCs also listed the fact that "a lot of change is required for PCMH" as a main challenge. Approaches to this included "standardize, routinize, change processes so they are more predictable," "hire [more] back-office staff for chart review, documentation, so the doctor can focus on patient main visit," and to gain "organizational commitment from senior leadership to the staff."

EHR Implementation

EHR implementation in general was also raised as a challenge. In our interview sample, demonstration respondents acknowledged how specific EHR features facilitated PCMH transformation and the recognition process through better documentation and tracking of care, as well as better coordination of care and population management. But many sites found that substantial modifications to their EHR functionality and usage were required to support PCMH transformation and recognition, such as new templates for documenting care, standardizing data input (including training staff to input data consistently), and new reporting and output functionality (e.g., care plans, NCQA documentation). In addition, EHR implementation in a number of sites occurred concurrently with participation in the FQHC APCP demonstration, in which case, EHR implementation typically took priority over PCMH efforts until the EHR system was functional.

Problems with EHR systems were also reported for specific components of the PCMH model, including poor usability, time-consuming data entry, interference with face-to-face patient care, inability to exchange health information with EHR systems of other facilities and providers, and degradation of clinical documentation.

Documentation

The problem of documentation in general for the NCQA application process was mentioned as an overarching concern for a majority of the FQHC respondents. Not only was there a concern about documenting the specifics of the six NCQA PCMH standards, the respondents also described the demanding nature of documenting every task that a clinician or provider engages in during a patient encounter.

> "The biggest thing is the quality of things that NCQA looks at and documenting each thing . . . in terms of education of patients . . . Are we giving them handouts? Are we really talking to them about their medications? Are we even checking to see whether or not they understand what we're saying to them? And these are things that we think we do? And, again, when I say we, I'm talking about providers asking and documenting these things."

> "With the NCQA, it's all on the documentation. So, if you have left out something that they were really looking for in the documentation, you're not going to pass that element or factor. And that's really, really difficult, telling this story strictly on paper."

One PCMH team at a site approached the documentation of the NCQA application process successfully by working iteratively with providers—providing "options" for how to document particular standards of care, "hear what their side is," and "give feedback in a nice way"—so that they "now have very, very comprehensive progress notes" documented in a way that is meaningful to providers and capable of being consistently reported out of the EHR system.

The majority of demonstration respondents we interviewed reported that the NCQA application process in general is costly and requires lots of documentation, especially for the patient education and coordination of care processes. Several respondents also indicated that PCMH demands large changes in staffing models to be sustainable and systemwide, especially as they tried to move from NCQA Level 2 to Level 3.

A few FQHCs mentioned the difficulties they were having with their patient experience survey tools as they moved from a few questions to a more standardized set of specific patient experience questions, such as using the Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey.

Lack of Site Interest in TA Offered by PCAs

Several PCAs discussed in detail the challenge stemming from a proportion of the demonstration sites within their state that are not interested in receiving TA from the PCMH. Possible reasons offered by PCA leaders included that sites believed they were progressing on their own without help, a general inclination—or even strategy—to wait

until closer to the October 2014 recognition deadline before fully engaging in the PCMH effort, and barriers of distance or independent-mindedness.

Yet there was also a sense that these uninterested sites would become more inclined to engage with PCAs as the final year of the demonstration approaches and the sites begin to realize the complexity of the PCMH change and recognition processes.

Unintended Consequences of the PCMH Recognition Process

Another challenge, related more to providing TA than engaging sites per se, concerned an unintended effect of the recognition process in focusing attention of sites (and PCAs) on the process of documentation over practice transformation.

> "The only other one I would offer is the unintended consequence of the very nature of the PCMH recognition application, which is to say, we built our coaching program and our technical assistance program on supporting transformative change and helping the health centers to change their health care delivery systems. But what happens is that, rather than rising to that, we sink a step or two below and we're simply focusing on documentation, on getting those points, on creating a policy that gets us this point. But it's hard to then translate that into true transformative change, and so there's a disconnect. That's a challenge, because the coaches really want to help our health centers move the needle, and instead they're proofreading policy. I know that's not what NCQA or any other accrediting body wants, but it is just the nature of the work."

This was viewed as especially problematic for sites that have assigned the PCMH effort to single individuals as a side project, "an NP [nurse practitioner] or a QI [quality improvement] coordinator or even an MA [medical assistant] who's kind of leading this charge . . . and . . . successful in getting the day-to-day documentation work done"—which characterizes the majority of sites. This was considered less of an issue for the "very small set of health centers with a long-term culture to be very forward-thinking [for whom] something like PCMH is very natural . . . and the change truly is happening."

Multiple Competing Priorities and Activities of Sites

PCA leaders also noted the multiple priorities and improvement activities vying for the attention of FQHCs, including the demonstration sites, which limited the ability of many sites to focus sufficient attention on the demonstration, as well as to coordinate participation across sites. Regarding the latter, one PCA leader wondered whether web or Internet tools might be used to reduce some of the less necessary real-time interactions.

Developing Relationships and Familiarity with Sites

PCA leaders also emphasized that many demonstrations are doing well when engaged. Although some may have enough resources and know-how on their own, others have shown consistent participation, been very responsive and held "two-way conversations" with the PCA, and have progressed to the point where they may require briefer, less-intensive interactions with the PCA to keep them on track.

However, it was also noted that achieving that level of participation and engagement is often initially dependent on building a relationship and level of familiarity between the sites and the PCA. Sites that have a relationship with the PCA were considered more likely to engage in the PCA's TA programs, and greater familiarity allows PCA staff to engage sites at a higher level. Another PCA leader described their experience that engaging sites was easier in smaller groups.

Reduced PCA Capacity Due to Staff Turnover

Two PCAs cited internal issues with insufficient capacity to provide TA due to turnover among PCA staff. One PCA that serves as a regional cluster lead mentioned that when this occurred in another state within its region, staff from one site were filling in at another site. Another PCA reported that it "had turnover, so capacity's been difficult internally," which created a situation where "I had to slow down on a lot of the work that I was doing. But we're slowly going to pick back up."

RQ1.D4 Why Do Some FQHCs Not Achieve Level 3 Recognition?

While this study question cannot be addressed until the end of the demonstration, experience from other medical home pilots (especially the TransforMed National Demonstration Project) indicates that several baseline attributes of FQHC sites may be associated with failure to achieve the demonstration's goals for practice transformation. In particular, disadvantaged starting positions for site culture (e.g., lack of values alignment between site leaders and staff) and workplace environment (e.g., dissatisfied or burned-out staff) may predispose sites to fail.

CASE survey data show that while nearly 80 percent of respondents reported high job satisfaction, (no association with RAS level), approximately 30 percent reported moderate or great likelihood of leaving their sites within two years (nonsignificant trend toward lower likelihood at RAS Level 3), and approximately 30 percent reported being burned out (no association with RAS level). Approximately 36 percent reported chaotic practice atmospheres. Chaos was not associated with RAS levels, but respondents in rural sites were statistically significantly less likely than those in metropolitan sites to report

chaotic atmospheres. Future analyses will examine whether clinics whose clinicians and staff describe chaos are less likely to show evidence for transformation.

The CASE survey also assessed many aspects of practice culture, predominantly using scales from preexisting survey instruments. In general, these scale scores have no inherent meaning in an absolute sense; they are useful only to identify groups of sites and types of staff with lower and higher values on each measured dimension for later longitudinal analysis after the conclusion of the demonstration. At baseline, we present cross-sectional comparisons between clinicians (physicians, nurse practitioners, and physician assistants) and other staff and across RAS levels (stratified by clinicians and other staff categories). On each dimension presented in this section, higher values are considered "better" (that is, a higher score on the "stress scale" is coded so that a higher score represents lower levels of stress).

Scores on scales for stress, work control, team structure, situation monitoring, mutual support, relationship infrastructure, facilitative leadership, sensemaking, teamwork, work environment, culture of learning, adaptive reserve, and values alignment with leadership were not associated with RAS levels collected just prior to the fielding of the baseline CASE data.

We do not know from these analyses whether the lack of relationship between early RAS and early CASE reflects a problem with the constructs associated with the measurement tools, a problem in the reliability of CASE informants representing their FQHC site, or inadequate time for transformation to occur and be reflected with the measurement tools. In our next report, we will supplement these analyses with additional ones examining the same respondent at two points in time. We anticipate these analyses will help us understand which of these many CASE domains predict transformation as measured by high RAS scores or the corollary achievement of NCQA Level 3 recognition.

There were differences between clinicians and other staff on many of these scales. Of note, clinicians were less likely (with borderline statistical significance) in RAS Level 2 and Level 3 sites than in Level 1 sites to report having 75 percent or more of the time they needed to perform physicals for new patients. In other words, there was greater time pressure at RAS Level 2 and Level 3 sites than at RAS Level 1 sites. Similar patterns were present for time pressure in routine visits with established patients and urgent care visits, although these trends were not statistically significant.

RQ1.D5 How Else Does TA Help the Transformation Process?

In the site leader interviews, we also asked FQHC representatives about strategies that facilitate sites' ability to be responsive to the six major categories of the 2011 NCQA

standards for PCMH recognition. Sites able to utilize these facilitators reported they were making progress toward achieving their desired NCQA recognition levels.

Facilitating PCMH Standard 1: *Enhance Access and Continuity*

Demonstration respondents mentioned a variety of interventions that support teamwork and communication with the goal of enhancing access and continuity of care. A few respondents discussed how having collocated teams contributed to building trusting relationships and strong communication among teams. A few other respondents mentioned that new locations with dedicated office and meeting spaces enhanced team members' ability to work together and conduct team huddles. Another demonstration respondent emphasized the benefit of convening regularly scheduled team meetings involving all staff, which increased continuity and communication across provider teams.

> "The interaction, the communication, was just brought to a whole new level because the team sat together. So if someone was talking about a specific patient's needs or whatever that was, the rest of the team might have picked up that they were talking to such-and-such patient, they could then add whatever they might know about that patient, maybe that behavioral health person knows something to add to the care."

Facilitating PCMH Standard 2: *Identify and Manage Patient Populations*

When asked about anything that helped in their PCMH implementation, several demonstration respondents mentioned investing in additional EHR functionality for population management. As described at one demonstration site:

> "An additional change that we made, and really also driven by the medical home model, is that we also invested in a population management system that integrates with our electronic health record. And we did that because there was a realization that even though you pull a lot of data out of the electronic health record, it's not always actionable the way that it comes out. And so, with this population management system we put in place, we're now able to really focus in on clinical conditions, or focus in on certain populations and segregate out populations, and so that's been very exciting for the organization. It's been very exciting for our providers. It's made the medical home model, perhaps, make more sense, because we might have talked about population management or care coordination within the medical home model and certainly care coordination is addressed."

Facilitating PCMH Standard 3: *Plan and Manage Care*

Demonstration respondents in our baseline interview sample did not mention any specific facilitators related to planning and managing care.

Facilitating PCMH Standard 4: *Providing Self-Care and Community Support*

Demonstration respondents in our baseline interview sample did not mention any specific facilitators related to providing self-care and community support.

Facilitating PCMH Standard 5: *Track and Coordinate Care*

Next to access and continuity (PCMH Standard 1), PCMH Standard 5—tracking and coordinating care—was the area with the most discussion of interventions geared to helping sites overcome challenges. A few respondents pointed to facilitators relating to customized information technology solutions for seamlessly sharing discharge summaries, referral information, or details of specialist visits. Other respondents discussed how data-sharing with providers of patient charts and encounters in the ED greatly facilitated care coordination. As one described it, "coordinating is no longer a paper trail; it is automatically seen within the patient chart."

Other facilitators mentioned by demonstration respondents included hiring a referral coordinator to be a team member and signing inpatient agreements with hospitalist groups as facilitators to care coordination.

Based on the current analysis, TA on Standard 5 can be of great help to sites in providing strategies, templates, and models on several issues:

- strategies for building relationships and engaging hospital and specialty providers in a site's local area, including distinguishing strategies for different types of providers (e.g., hospitalists, care managers, behavioral health agencies), and for different types of local health systems, particularly for sites with patients utilizing multiple hospital and specialty providers (e.g., how to coordinate or consolidate these wider referral networks).
- templates and strategies for customized information technology solutions to facilitate managing the referral process internally and with external providers, particularly for managing receipt of follow-up information available from external providers, and for understanding the issues and costs in maintaining electronic systems to exchange referral information.
- templates and strategies for efficiently managing referral processes and information when information technology solutions are not available
- models of assigning responsibility of the referral coordination function to specific staff (either full-time referral coordinators or as part of their roles) and incorporating these individuals as part of the clinical care team
- strategies for integrating referral tracking and flows of information into routine process of care rather than perceived as an add-on, overly burdensome, paper-trail process

- strategies for meeting needs of patients when necessary services are not readily available or accessible (linguistically, transportation, etc.) in the local area.

Facilitating PCMH Standard 6: *Measure and Improve Performance*

Several demonstration respondents mentioned the importance of using a PCMH survey for their patient-experience surveys.

> "We've been doing patient surveys forever. We just switched to the patient-centered survey last year. So, you have some historical data you can compare things to . . . We'll be able to tell, for example, if the portal improves communication. Hopefully, that'll be reflected in the patient satisfaction survey. It should help us pinpoint changes and successes."

RQ1.D6 How Do the Interventions Help Sites Overcome Challenges?

Access to specialists and mental health services may challenge FQHC sites' transformation into APCPs. The CASE survey queried respondents about whether their sites were attempting to increase the amount of care that patients received from specialists and the availability of mental health services. While neither of these items completely addresses the challenges identified in RQ1.D3, they may help distinguish sites that are attempting to overcome challenges from those that are not.

Baseline CASE analyses show that 23 percent of respondents reported that their sites were making efforts to increase the amount of care their patients receive from specialists, and approximately 60 percent were making efforts to increase their patients' access to mental health services. These efforts were not associated with RAS level.

RQ1.D7 Which Intervention Features Are Most Useful and Which Intervention Features Are Not as Helpful?

PCAs emphasized thoughtful recommendations for those intervention features they believed most useful.

Developing Relationships and Familiarity with Sites

PCAs noted that achieving the desired level of participation and engagement often depends upon early development of a trusting relationship and level of familiarity between the sites and the PCAs. Sites that develop this type of relationship with their PCA were considered more likely to engage in the PCA's TA programs. That greater engagement led to more familiarity that, in turn, allowed PCA staff to engage sites at a higher level. Another PCA leader shared the experience that engaging sites was easier in smaller groups.

PCA leaders emphasized the value of developing direct relationships with site leaders and staff. Within the context of this relationship, PCA leaders could share their deep familiarity with the issues close to FQHCs and personalize the lessons learned to individual FQHCs.

> "Having that relationship with the site helped them to understand what backup I had, and I knew how their system works and we could troubleshoot areas . . . That kept them engaged at a higher level than some of the other health centers."

> "So with those who have been more engaged, it's much more of a two-way conversation. . ."

Mindfulness on the Dual Objectives of the Demonstration

To mitigate the chance that the PCMH recognition process would eclipse practice transformation, PCA leads recommended training with coaches to first address site-specific questions on recognition documentation, and to follow that with explicit discussions about practice changes required to implement and sustain required policies and procedures.

> "The coaches have all been trained that whatever question they may get, which is typically around, 'does this policy meet this standard,' the follow-up question is always, 'yes or no, it would or would not meet the standard.'"

> "But the real question we have to address is 'how do you really intend to implement it and what change do you have to support to ensure that this is something that you can sustain long-term for the benefit of the patient?' "

Tailoring TA/Integrating PCMH Efforts into the Site Quality Strategies

Building upon site-specific perspectives, PCAs can contextualize PCMH processes and provoke a focused dialogue with sites at a less formal level. This can provide opportunities to embed PCMH efforts into existing quality strategies so they are part of, and not extraneous to, existing work flows. This strategy of integrating with existing staff can also be valuable when leadership and staff are less than ideal for responding to the many demands of both transformation and recognition.

> "The struggle that I'm having is . . . mostly competing priorities at the health centers . . . there are other things going on. It's a lot of work and most health centers don't have a team of people working on this. You have just one person or two people working on it. So, the struggle that I have is really saying 'we all know that this is important, but what is that timeline that you're gonna set to get the work done?'"

Better Use of Web or Internet Tools

In response to staff limitations in sites, PCAs emphasized the potential for more frequent reliance on high-performing information technology tools. They emphasized that using these tools could reduce staff time required in less-crucial real-time interactions.

> "We get a big turnout, but we don't get everybody, and that's just never going to happen. That is the challenge with these things. And frankly . . . if we can do all this through online and web, that's the way to get the data done and all this other stuff, because trying to meet phone call meetings is really, really, really a challenge."

In contrast, interventions that did not seem helpful focused on tasks that distracted PCAs or sites away from patient care or away from transformation activities. PCAs expressed concerns that too many meetings, too many webinars, and too many conference calls were counterproductive. Conflicting priorities for focusing attention on NCQA application processes were not productive. Finally, the long delays experienced by some sites as they awaited a response about their NCQA recognition status were also considered not helpful.

Some Concluding Observations

Although FQHCs were enthusiastic about the hoped-for transformation to APCPs, many were overwhelmed with the need to change so much so fast while simultaneously completing complex forms required for NCQA recognition applications and still caring for complex patients. Many forces pressured participating FQHCs to respond to cues for prompt change, but for some, the time pressure compromised efficiency. Ongoing requests for sites to participate in webinars, calls, surveys, and site visits occurred concomitant with the need to respond to six domains of transformation heralded by NCQA and its associated RAS. At the same time, sites were implementing new her systems, developing new forms of cultural competence, and attempting to improve patient access, practice accountability, and teamwork. Some sites described teams that were exhausted and frustrated, others described frenzy, apathy, or a sense of having gone astray.

While we report on progress made during the first two years of the demonstration, our qualitative and quantitative findings illustrate that demonstration sites experienced many barriers to PCMH change. Some difficulties were associated with generic quality improvement efforts; others were associated with challenges unique to implementing the PCMH model of care. Prevalent themes pertinent to barriers to PCMH transformation include:

- *Provider and staff resistance* to PCMH change in many FQHCs related less to objections to PCMH concepts (which are typically viewed as compatible with the mission of community health centers) and more to a general reluctance to change, especially given the competing priorities to which FQHCs and their staff must attend.
- The *extent of change and effort required* for PCMH transformation and recognition is also often underestimated by FQHC leaders and other staff.
- *Educating leaders and staff about the process of PCMH change and recognition* was a challenge and priority in many sites.
- *Framing the value of best practices in use at the FQHC* is challenging because providers and other staff often perceive systematizing these practices as being associated with increased workload.
- *Integrating and embedding new PCMH-related tasks into routine processes* is complex because it involves alignment of the full set of team members, the EHR system, and the patients. Nevertheless, this is needed to ensure that providers and staff experience changes as improving—rather than adding to—their workflow and burden.
- *Documentation of care policies and processes*, both for NCQA recognition and for implementing and ongoing monitoring of PCMH changes, was also a widely reported challenge. This typically required iterative interactions with providers and updating of data input practices and reporting functionality, even in sites with developed EHR systems.
- *EHR functionality in general* was also raised as a challenge, given the centrality of EHR systems to implementing PCMH changes. In sites that implemented EHR and PCMH concurrently, EHR implementation typically took priority; many sites also reported difficulties working with EHR vendors and information technology support staff to tailor systems to PCMH needs.
- *Access to external specialists* was reported as a major challenge to fully implementing a PCMH model of care, even though NCQA standards focus specifically on coordination and tracking of specialty referrals once established.
- *Lack of site interest in direct TA* was reported by several PCAs as well as sites. This was often reflected in sites postponing PCMH efforts until closer to the demonstration deadline, either because they felt they were making sufficient progress on their own, or because they faced barriers of geographic distance or independent-mindedness. PCAs tended to address this challenge through strategies to build relationships and familiarity with sites.

Many sites, especially those with an advanced starting position with respect to APC principles, thrived in the demonstration. But others, regardless of their starting position, fumbled with too little time and too little guidance. Leadership with a commitment to APC principles emerged as a key component of success, especially when the leadership was continuous and resilient. Unfortunately, that leadership was not consistently available for a substantial number of sites.

III. Key Policy Question 2: Do Demonstration Sites Deliver Better Beneficiary Processes and Outcomes Than Comparison Sites?

Exhibit III.1: 13 RQs and their Derivative Subquestions for Key Policy Question 2

Second Annual Report Section	Research Question
III.RQ2.1	Do FQHCs participating in the demonstration provide improved loyalty or continuity of care for Medicare beneficiaries?
III.RQ2.2	Do FQHCs participating in the demonstration provide more timely delivery of health services to Medicare beneficiaries?
III.RQ2.3	Do FQHCs participating in the demonstration provide better or enhanced access to Medicare beneficiaries' PCMH providers?
III.RQ2.4	Do FQHCs participating in the demonstration provide improved adherence to evidence-based guidelines?
III.RQ2.5	Do FQHCs participating in the demonstration generate better self-rated quality of care for Medicare beneficiaries?
III.RQ2.6	Are Medicare beneficiaries, their family members, and/or their caregivers served by participating FQHCs able to participate more effectively in decisions concerning their care?
III.RQ2.7	Are Medicare beneficiaries served by the participating FQHCs better able to self-manage their health conditions or more likely to engage in healthy behaviors?
III.RQ2.8	Do FQHCs participating in the demonstration provide better experiences with the health care system for Medicare beneficiaries and their families and caregivers?
III.RQ2.9	How does FQHC participation in the demonstration affect utilization of services covered by Medicare?
III.RQ2.10	How does FQHC participation in the demonstration affect expenditures—both out-of-pocket costs (deductibles and coinsurance payments) and costs covered out of the Medicare Trust Fund—for services covered by Medicare?
III.RQ2.11	Do FQHCs participating in the demonstration provide better coordination of care for Medicare beneficiaries?
III.RQ2.12	Do FQHCs participating in the demonstration provide reductions in, or elimination of, health care disparities among Medicare beneficiaries?
III.RQ2.13	Does the demonstration cause any spillover effects on Medicaid beneficiaries or other populations served by the FQHCs?

Appendix R provides a cross walk between RQs discussed in this section and those that have previously been linked to Key Policy Question 2.

Section III is divided into three sections. First, we present a summary of the approach to and findings for Key Policy Question 2. Second, we use Medicare beneficiary data and beneficiary survey data to individually introduce beneficiary-level analyses comparing users of demonstration FQHCs with users of comparison FQHCs for 12 RQs. Finally, for

RQ13, we use Medicaid beneficiary data to introduce beneficiary-level analyses of Medicaid beneficiaries associated with demonstration and comparisons sites.

Summary of Approach to and Findings for Key Policy Question 2

Key Policy Question 2 asks whether Medicare beneficiaries assigned to demonstration sites experience better processes and outcomes of care than those experienced by beneficiaries at comparison sites. To assess the effectiveness of the APCP demonstration, Key Policy Question 2 focuses on differences between demonstration and comparison FQHC sites that may lead to better management of chronic conditions, improved health outcomes, decreased use of inappropriate health care services (e.g., certain hospitalizations, ED visits, duplicative or unnecessary tests and procedures), increased use of other appropriate services (e.g., preventive services), and reductions in health care expenditures.

To answer this policy question, the evaluation focuses on metrics spanning 12 RQs pertinent to APCP principles, including: (1) continuity, (2) timeliness, (3) access to care, (4) adherence to evidence-based guidelines, (5) beneficiary ratings of providers, (6) effective beneficiary participation in decision making, (7) self-management, (8) patient experiences with care, (9) coordination of care, (10) disparities, (11) utilization, and (12) expenditures. Some of these metrics are evaluated using claims data, others by beneficiary report. Both data sources are needed to meaningfully understand whether demonstration interventions designed to stimulate APC principles are associated with improvements in processes and outcomes of care for Medicare beneficiaries. Claims analyses provide the advantage of a very large sample with a longitudinal data analysis; beneficiary survey analyses provide rich clinical and patient experiential data. Also included in our evaluation of Key Policy Question 2 are data from the CASE survey and qualitative analyses that provide context and nuance for interpreting all the available data.[20]

Available results for Key Policy Question 2 include analyses of differences between demonstration and comparison sites for all 12 RQs spanning 58 different metrics.[21] The early beneficiary survey, collected midway through the three-year demonstration,

[20] By design, these follow-up primary data collection efforts were planned toward the end of the demonstration so that our longitudinal analyses could maximize the difference between early (baseline) and late (follow-up) experiences. The final report will include longitudinal analyses of all three of these primary data sources as a supplement to a more complete set of claims data.

[21] Significant differences are defined as $p<0.05$ for differences between demonstration and comparison sites.

provides data for 46 metrics spanning all 12 RQs. Claims data contribute to four of the 12 RQs spanning 12 metrics.

Of the 46 beneficiary survey–based metrics, 85 percent of the analyses comparing demonstration FQHCs with comparison FQHCs showed no significant difference at this stage of the analysis. Further beneficiary survey analyses will be included in the final report after longitudinal data become available.

Claims-Based Analyses

Claims-based analyses pertinent to utilization, cost, evidence-based care, and continuity of care are summarized here. After selecting claims-based evaluation metrics for these RQs, we assessed the impact of the demonstration on evaluation metrics by comparing changes over time within the FQHC demonstration cohort relative to changes over time within the comparison cohort for the original intention-to-treat (ITT) cohort.

Claims-Based Utilization and Cost Findings

Across nine quarters comparing demonstration FQHCs with comparison FQHCs adjusted for differences at baseline, we noted significantly more utilization for three of four claims-based utilization measures and more total costs for the demonstration users than for the comparison users. Specifically, in contrast with comparison FQHC users, demonstration FQHC users had significantly more use of

- hospital admissions in two of nine demonstration quarters
- admissions for ambulatory care sensitive conditions (ACSCs) in none of nine demonstration quarters
- unplanned hospital readmissions in one of nine demonstration quarters
- ED visits in six of nine demonstration quarters
- total Medicare payments in four of nine quarters.

There was no demonstration quarter in which demonstration site users had significantly less utilization or less total cost than comparison sites.

Claims-Based Evidence-Based Metric Findings

Across nine quarters comparing demonstration FQHCs with comparison FQHCs relative to differences at baseline, the demonstration group significantly outperformed the comparison group for at least eight quarters for glycated hemoglobin blood tests (HbA1C), retinal eye exams, and nephropathy testing, though we observed little difference for low-density lipoprotein (LDL) testing for diabetic patients or for lipid testing for ischemic vascular disease patients.

Claims-Based Continuity of Care Findings

Adjusted analyses show no statistically significant differences in the Bice-Boxerman continuity of care measure between the demonstration and comparison FQHC groups in the first year of the demonstration compared with baseline. In year 2, demonstration FQHCs are associated with a statistically significant but very small (~1 percent) decrease in continuity when looking across all primary care provider visits and when looking at primary care and specialist care together.

Beneficiary Survey–Based Analyses

Among the primary beneficiary survey–based analyses, demonstration FQHCs perform better than comparison FQHCs for six of 46 (13 percent) beneficiary survey–based metrics including

- more frequent beneficiary interviews regarding mental health needs*[22]
- better provider support for beneficiary self-care*
- more frequent specialist visits*
- more frequent use of mental health services*
- specialists' knowing beneficiaries' most important medical history
- post-hospital discharge visits within 14 days of discharge.*

However, the significance of all but one of these differences is lost after correction for multiple comparisons.

Within the primary analyses, demonstration FQHCs perform worse than comparison FQHCs only for beneficiary report of use of dental services within the last 12 months.

Summary of Process and Outcomes Differences Between Demonstration and Comparison Groups

Exhibit III.2 provides a summary of all 58 metrics spanning 12 RQs. Column 1 lists whether the metric is derived from the beneficiary survey or claims data. In column 2, the light gray rows specify the topic of the RQ and the white rows provide a brief description of the metrics. The third column summarizes results of analyses comparing demonstration and comparison FQHCs.

[22] An asterisk indicates that adjustment for multiple comparisons with the Bonferroni correction is associated with loss of significance for this variable. Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association 56*(293): 52–64.

Exhibit III.2: Metric Performance for Demonstration FQHCs Relative to Comparison FQHC: Metric-Level Analyses

The following tables show differences for demonstration FQHCs relative to comparison FQHCs.

RQ2.1 Improved loyalty or continuity of care (No difference)

Data Source	Metrics	Differences
Bene survey 1	Beneficiary continuity >= 5 years with a particular provider	No difference
Bene survey 2	Beneficiary value with the roles fulfilled by their attributed providers	No difference
Bene survey 3	Beneficiaries seeing their own personal doctor or nurse compared with a different provider	No difference
Claims 4	Loyalty based on attribution using claims	No difference
Claims 5	Continuity of care using claims	No difference

RQ2.2 More timely delivery of health services to Medicare beneficiaries (No difference)

Data Source	Metrics	Differences
Bene survey 6	Always received an appointment as soon as needed when made an appointment for check-up or routine care	No difference
Bene survey 7	Always received an appointment as soon as needed when phoned to get an appointment for care right away	No difference
Bene survey 8	Always received answer to a medical question posed during regular office hours on that same day	No difference
Bene survey 9	Always received an answer to a medical question that was posed by phone after regular office hours, as soon as needed	No difference
Bene survey 10	Always saw provider within 15 minutes of their appointment time	No difference
Bene survey 11	Always getting the care needed during evenings, weekends, or holidays if phones for care needed right away	No difference
Bene survey 12	Days waiting for an appointment when needing care right away	No difference

RQ2.3 Better access (No difference)

Data Source	Metrics	Differences
Bene survey 13	Access to information about what to do if care needed after hours	No difference
Bene survey 14	Access to information about services that occur between visits	No difference
Bene survey 15	Access to specialists	No difference
Bene survey 16	Access to needed help with transportation provided to satisfaction	No difference
Bene survey 17	Satisfaction with interpreter services	No difference

RQ2.4 Improved adherence to evidence-based care (No difference or demo more likely or increasing)

Data Source	Metrics	Differences
Bene survey 18	Immunization evidence-based guidelines	No difference
Bene survey 19	Primary prevention of cardiovascular disease with aspirin use	No difference
Bene survey 20	Colorectal cancer screening	No difference

Data Source	Metrics	Differences
Bene survey 21	Engaging beneficiaries who smoke in discussions about evidence-based smoking cessation interventions	No difference
Bene survey 22	Counseling regarding weight loss, exercise and eating right	No difference
Bene survey 23	Mental health interventions	Demo more likely*
Claims 24	HbA1c testing for diabetic patients from claims	Demo increasing
Claims 25	Eye examination for diabetic patients from claims	Demo increasing
Claims 26	Nephropathy testing for diabetic patients from claims	Demo increasing
Claims 27	LDL testing for diabetic patients from claims	Demo the same or increasing
Claims 28	Lipid testing for patients with ischemic vascular disease patients from claims	No difference

RQ2.5 Better quality of care to Medicare beneficiaries (No difference)

Data Source	Metrics	Differences
Bene survey 29	Beneficiary ratings of their attributed providers	No difference
Bene survey 30	Beneficiary ratings of specialty providers	No difference
Bene survey 31	Beneficiary ratings of “helpful, courteous and respectful staff”	No difference

RQ2.6 More effective participation in health care decisions (No difference)

Data Source	Metrics	Differences
Bene survey 32	Effective participation in decisions about medication use	No difference

RQ2.7 Better able to self-manage (No difference or demo more likely)

Data Source	Metrics	Differences
Bene survey 33	Educating beneficiaries for self-management	No difference
Bene survey 34	Provider support for beneficiaries caring for their own health	Demo more likely*
Bene survey 35	Provider follow-up on test results	No difference

RQ2.8 Have better experiences with the health care system (No difference)

Data Source	Metrics	Differences
Bene survey 36	Beneficiary experiences	No difference

RQ2.9 Change utilization of services (No difference or demo more likely or increasing)

Data Source	Metrics	Differences
Bene survey 37	Provider visits and tests	No difference
Bene survey 38	Hospitalizations	No difference
Bene survey 39	Vision services for adults	No difference
Bene survey 40	Vision services for diabetic adults	No difference
Bene survey 41	Services for impaired hearing	No difference
Bene survey 42	Receipt of blood or x-ray test	No difference
Bene survey 43	Use of prescription medications	No difference
Bene survey 44	Use of specialist visits	Demo more likely*
Bene survey 45	Use of mental health specialty services at least once within the last 12 months	Demo more likely*

Data Source	Metrics	Differences
Bene survey 46	Use of dental services at least once within the last 12 months	Demo less likely*
Claims 47	Inpatient admissions from claims	Demo same or increasing
Claims 48	Admissions rate for chronic ACSCs from claims	No difference
Claims 49	Unplanned hospital readmissions from claims	Demo same or increasing
Claims 50	ED visits from claims	Demo same or increasing

RQ2.10 Affect expenditures (No difference or demo same or increasing)

Data Source	Metrics	Differences
Bene survey 51	Beneficiary awareness of costs of seeing a specialist	No difference
Claims 52	Total Medicare payments from claims	Demo same or increasing

RQ2.11 Provide better coordination of care (No difference or demo more likely)

Data Source	Metrics	Differences
Bene survey 53	Specialist knows most important medical history	Demo more likely
Bene survey 54	Visiting or calling a clinic within 14 days following hospitalization	Demo more likely*
Bene survey 55	Coordination from the attributed provider's position	No difference
Bene survey 56	Discussing and helping with home health services	No difference

RQ2.12 Provide reductions in health care disparities (No difference)

Data Source	Metrics	Differences
Bene survey 57	Unfair treatment because of race or ethnicity	No difference
Bene survey 58	Unfair treatment because of English language skills	No difference

* Within these exhibits, an asterisk within the cell indicates that adjustment for multiple comparisons is associated with loss of significance for this variable.
[a] Differences refer to differences across the demonstration quarters between demonstration and comparison sites as compared with differences during the 12-month baseline period that immediately preceded the demonstration.

In summary, our analyses to date show significant differences between demonstration and comparison FQHCs for most measured claims-based metrics, but no differences for most beneficiary survey–based metrics.

Claims-based longitudinal analyses across nine quarters show significantly more utilization and costs for demonstration FQHCs than comparison FQHCs for hospital admissions (two quarters), readmissions (one quarter), ED visits (six quarters), and total Medicare payments (four quarters).

Beneficiary survey-based cross-sectional analyses show few differences, and several of these lose significance after adjustment for multiple comparisons. Beneficiary survey–based analyses will be expanded into longitudinal analyses in our next report.

Our work during the next year will include analyses of additional claims data, the completion of analyses of longitudinal beneficiary survey responses, and analyses of clinician and staff experience surveys and qualitative data. Though the impact of the effect of the demonstration on total costs is not likely to reverse (from our current reports of demonstration sites costing more than comparison sites), the next steps in the analyses will examine whether beneficiary experiences change as the demonstration is completed.

Analyses of 12 Individual Research Questions Pertinent to Key Policy Question 2

This section introduces our approach to each of the 12 RQs that support Key Policy Question 2. For each RQ, we list the data sources, summary of methods, and results. More detailed methods associated with each of these data sources are documented in Appendixes B (claims methods), C (beneficiary survey), D (RAS/NCQA data), E (CASE methods), and F (qualitative methods).

Each individual RQ is assigned a number (i.e., RQ1 through RQ12). We describe methods and findings associated with each RQ in consecutive order as illustrated in Exhibits III.1 and III.2. We begin our discussion of each RQ with a statement of the question followed by a brief summary of methods and findings. Many of the RQs are evaluated using more than one data source (including beneficiary survey, claims, CASE survey, RAS/NCQA data, or qualitative analyses). Additionally, within or across data sources, more than one metric may be used to address the RQ. Across the 12 RQs, we highlight 58 metrics defined using either beneficiary survey or claims data, though CASE, NCQA/RAS, or qualitative data are often presented to contextualize findings. For each summary, we select the level of aggregation that is likely to help the reader best understand the analytic findings to date. Text describing each RQ may span several pages to allow the reader to understand the basis for our analytic findings.

To highlight the most pertinent information across the many RQs, in this section we present data tables for only a limited number of detailed analyses. When beneficiary SQ analyses are used to answer RQs, we orient the reader by naming the individual or group of analyzed survey items within a light gray bar that is followed by three pieces of information. First, we list the beneficiary SQ number (in brackets) followed by a description of the beneficiary cohort for whom the presented data are applicable. Next, we show the significance p-values for demonstration versus comparison FQHCs. Displayed p-values are calculated across all response categories unless the text only references the “always” category, in which case that is the only category assessed. To

help the reader see the large number of *not statistically significant* baseline beneficiary survey comparisons, we italicize these not-significant values in light gray.

The RQ analyses that follow assess differences between demonstration and comparison FQHC users. CASE and qualitative analyses provide contextual information where relevant.

For many of the 12 RQs shown in Exhibit III.1, there is substantial interest in two types of derivative questions. The first relates to the facilitators of improvement (or change) for the metrics listed, and the second pertains to the impact that changes in these metrics will have on key demonstration outcomes, such as utilization and cost. While Key Policy Question 2 focuses on differences between demonstration FQHC and comparison site users across 58 metrics, Key Policy Question 3 focuses on site and beneficiary characteristics that facilitate these differences or that influence the impact of these differences on outcomes such as cost and utilization. Proposed methodologies for these latter analyses are discussed in Section IV.RQ3.3 and Exhibit IV.3.

RQ2.1 Do FQHCs Participating in the Demonstration Provide Improved Loyalty or Continuity of Care for Medicare Beneficiaries?

RQ2.1.a Summary: Loyalty and Continuity of Care (Baseline beneficiary survey)

The beneficiary survey provided three measures of loyalty and continuity that allow us to evaluate whether demonstration FQHC users differ from comparison FQHC users.

The first measure is beneficiary continuity with a particular provider. Continuity is assessed as the length of time beneficiaries report having been a patient of their attributed provider.

- Demonstration FQHC users were not different from comparison FQHC users ($p=0.9269$).

The second measure of loyalty/continuity is the degree to which beneficiaries value their providers. For example, among beneficiaries who visited their attributed providers within the last 12 months, we assessed the prevalence of beneficiaries reporting their attributed provider was the most helpful to them for each of the following: (1) deciding whether to have tests or treatments; (2) helping the beneficiary change their health habits; (3) understanding how to approach their most important medical problem; (4) knowing their provider is the one in charge of following up on health and medical conditions if help is needed.

- Demonstration FQHC users were not different from comparison FQHC users ($p>0.05$).

A third measure of loyalty/continuity is assessed by beneficiaries seeing their own personal doctor or nurse compared with a different provider, among beneficiaries who visited their attributed offices.

- Demonstration FQHC users were not different from comparison FQHC users (p=0.4599).

RQ2.1.b Analysis: Loyalty and Continuity of Care (Baseline beneficiary survey)

Loyalty/continuity defined by the duration of time a beneficiary was a patient of the attributed provider
[Analysis SQ2.4C.3] Among beneficiary respondents who ever received care from their attributed clinic, • 50.80 percent of demonstration FQHC vs. 50.75 percent of comparison FQHC users reported having been a patient of their attributed provider *for at least five years* – p=0.9269 for demonstration FQHC vs. comparison FQHC.
Loyalty/continuity defined by beneficiaries recognizing value in their providers
[Analysis SQ2.4B.1] Among beneficiary respondents, • 88.15 percent of demonstration FQHC and 86.84 percent of comparison FQHC users reported that *their attributed provider is the one they usually see if they need a check-up, want advice about a health problem, or get sick or hurt* – p=0.1813 for demonstration FQHC vs. comparison FQHC.
[Analysis SQ2.4B.2] Among beneficiary respondents, • 86.99 percent of demonstration FQHC and 85.92 percent of comparison FQHC users reported that *the provider they saw on their most recent visit is the provider who has been most helpful during the last 12 months in helping them decide whether or not to have tests or treatments, or to change their health habits* – p=0.3167 for demonstration FQHC vs. comparison FQHC.
[Analysis SQ2.4B.4] Among beneficiary respondents, • 86.30 percent of demonstration FQHC and 85.24 percent of comparison FQHC users reported that *the provider they saw on their most recent visit is the provider who has been most likely to help them with their most important medical problems* – p=0.3258 for demonstration FQHC vs. comparison FQHC.
[Analysis SQ2.4B.5] Among beneficiary respondents, • 89.71 percent of demonstration FQHC and 89.26 percent of comparison FQHC users reported that *the provider they saw on their most recent visit is the provider who is in charge of following up on their health and medical conditions if they need help* – p=0.6230 for demonstration FQHC vs. comparison FQHC.
Loyalty defined by beneficiaries seeing their own personal doctor or nurse
[Analysis SQ2.4B.6] Among beneficiary respondents who visited their attributed clinic within the last 12 months, • 9.65 percent of demonstration FQHC and 7.10 percent of comparison FQHC users reported that *when*

they had a visit at their personal doctor or nurse's office in the last 12 months that they never saw their personal doctor or nurse, while 66.84 percent of demonstration FQHC and 66.33 percent of comparison FQHC users reported *always seeing their personal doctor or nurse*

- p=0.4599 for demonstration FQHC vs. comparison FQHC.

RQ2.1.c Summary: Loyalty and Continuity of Care (Claims data)

Loyalty Based on Attrition

Preliminary analyses examined the loyalty of beneficiaries attributed to a demonstration or comparison FQHC in the first quarter of the demonstration. In terms of attrition and switching to another source of usual care over the course of the demonstration, demonstration FQHC users were not different from comparison FQHC users.

These claims based-findings are consistent with the beneficiary survey analyses.

Continuity of Care Using Bice-Boxerman Index

We used the Bice-Boxerman continuity of care index to assess continuity of care at the patient-provider level across all of (1) primary care, (2) specialty care, and (3) primary and specialty care combined together. For these measures, demonstration FQHC users were not different from comparison FQHC users.

RQ2.1.d Analysis: Loyalty and Continuity of Care (Claims data)

Loyalty Based on Attrition

Preliminary analyses that have examined the loyalty of beneficiaries attributed to a demonstration or comparison FQHC in the first quarter of the demonstration show that there is not an obvious difference among groups in terms of attrition and switching to another source of usual care over the course of the demonstration.

Specifically, Exhibits III.3 and III.4 show that among beneficiaries attributed to a demonstration FQHC at the start of the demonstration, 58 percent remain attributed to the same FQHC at the end of demonstration Quarter 7, while 12 percent are attributed to another FQHC, PCC, or rural health center (RHC) at the end of this period. These rates are similar in the comparison group.

Continuity of Care Using Claims

Background

Among beneficiaries attributed to a demonstration or comparison FQHC during the first quarter of the demonstration (FQHC intention-to-treat [ITT] cohort), we assessed patient continuity of care with providers before and after the start of the demonstration

using the Bice-Boxerman index, which is a measure of patient-provider continuity that ranges from 0 to 1, with increasing values indicating greater continuity between a patient and a provider.[23] The measure incorporates the number of unique providers seen, the total number of visits with each provider, and the total number of visits to all providers over a defined period of time. We calculated this index for the baseline, first, and second years of the demonstration separately, using outpatient Medicare claims data.

For each of the three years, we examined continuity among (1) all primary care providers, including those seen within and outside of an FQHC, (2) all specialists, and (3) both primary care and specialists. Clinicians in internal medicine, general practice, family medicine, obstetrics and gynecology, adult health, community health, family practice, primary care, women's health, medical, gerontology, pediatrics, and preventive medicine, were classified as primary care providers. Specialist categories included cardiology, allergy and immunology, dermatology, emergency, endocrinology, ENT, optometry, gastroenterology, hematology and oncology, hospice, mental health, neurology, nephrology, orthopedics, surgery, urology and others. Primary care and specialist provider categories were defined by National Plan and Provider Enumeration System Provider Taxonomy Codes and limited to physicians, nurse practitioners, and physician assistants. Unique providers were identified using National Provider Identifier (NPI) codes.

The Bice-Boxerman index is uninformative for patients with few visits. Therefore, we excluded patients with fewer than three visits when calculating the index within each of the three provider categories (primary care, specialists, and both).

[23] Bice, T.W., Boxerman, S.B. (1977). A quantitative measure of continuity of care. *Medical Care, 15*(4), 347–349.

Exhibit III.3: Loyalty of Beneficiaries Attributed to Demonstration FQHCs in the First Quarter of the Demonstration

Attribution Quarter	Q5 Demo FQHC N (percent)	Other Demo FQHC N (percent)	Comp FQHC N (percent)	Other FQHC N (percent)	PCC N (percent)	Rural Health Center (RHC) N (percent)	Not Attributed N (percent)
1	152,308 (100)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
2	133,698 (87.8)	714 (0.5)	110 (0.1)	1,287 (0.8)	5,427 (3.6)	780 (0.5)	10,292 (6.8)
3	121,305 (79.6)	1,204 (0.8)	202 (0.1)	2,281 (1.5)	8,011 (5.3)	1,314 (0.9)	17,991 (11.8)
4	110,962 (72.9)	1589 (1)	280 (0.2)	2,977 (2)	8,844 (5.8)	1,747 (1.1)	25,909 (17)
5	101,574 (66.7)	2,055 (1.3)	345 (0.2)	3,554 (2.3)	8,474 (5.6)	2,144 (1.4)	34,162 (22.4)
6	93,809 (61.6)	2,266 (1.5)	367 (0.2)	3,853 (2.5)	8,107 (5.3)	2,368 (1.6)	41,538 (27.3)
7	87,951 (57.7)	2,378 (1.6)	391 (0.3)	4,086 (2.7)	7,867 (5.2)	2,544 (1.7)	47,091 (30.9)

Exhibit III.4: Loyalty of Beneficiaries Attributed to Comparison FQHCs in the First Quarter of the Demonstration

Attribution Quarter	Q5 Comp FQHC N (percent)	Other Comp FQHC N (percent)	Demo FQHC N (percent)	Other FQHC N (percent)	PCC N (percent)	RHC N (percent)	Not Attributed N (percent)
1	275,869 (100)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
2	242,229 (87.8)	2,260 (0.8)	148 (0.1)	1,442 (0.5)	9,355 (3.4)	2,266 (0.8)	18,169 (6.6)
3	219,218 (79.5)	3,728 (1.4)	242 (0.1)	2,507 (0.9)	13,957 (5.1)	4,448 (1.6)	31,769 (11.5)
4	199,798 (72.4)	4,761 (1.7)	327 (0.1)	3,275 (1.2)	15,656 (5.7)	6,031 (2.2)	46,021 (16.7)
5	182,136 (66)	5,792 (2.1)	401 (0.1)	3,946 (1.4)	15,149 (5.5)	6,761 (2.5)	61,684 (22.4)
6	166,992 (60.5)	6,164 (2.2)	427 (0.2)	4,141 (1.5)	14,715 (5.3)	7,094 (2.6)	76,336 (27.7)
7	155,931 (56.5)	6,319 (2.3)	456 (0.2)	4,200 (1.5)	14,329 (5.2)	7,302 (2.6)	87,332 (31.7)

Analysis Methods

We used a difference-in differences regression approach to determine whether patient-provider continuity of care differed between the demonstration and comparison groups after the start of the demonstration compared with how they differed before the start of the demonstration, for primary care, specialist care, and both. Models were fit using Generalized Estimating Equations (GEE)—generalized linear models with an identity link and a normal error distribution, clustered at the beneficiary level. Models examined whether there was a difference in provider continuity between demonstration and comparison groups in Year 1 of the demonstration compared with baseline, and in Year 2 of the demonstration compared with baseline, controlling for the number of annual provider visits, as well as the same beneficiary, site, grantee, and area-level characteristics included in the primary outcomes regression models. Propensity score weights were also used in conjunction with these models to ensure balance between demonstration and comparison groups. Propensity score models included all covariates included in the primary outcomes evaluation models plus baseline continuity of care scores.

Results

Unadjusted

Continuity of care measures for primary care visits, specialist visits, and both are shown for the baseline year and Years 1 and 2 of the demonstration in Exhibit III.5, for the demonstration and comparison groups separately. We see that the continuity of care index is consistently higher for primary care compared with specialty care for both the demonstration and comparison groups. Within primary care and looking at primary care together with specialists, continuity of care decreases over time for both the demonstration and comparison groups. Among specialists alone, continuity of care decreases in the first year of the demonstration compared with baseline and then increases in the second year, in both groups.

Exhibit III.5: Unadjusted Bice-Boxerman Index for Primary Care Visits, Specialist Visits, and Both, for Baseline Year and First Two Years of the Demonstration, Stratified by Demonstration and Comparison Groups

		Demonstration FQHCs		Comparison FQHCs	
	Year	N	Mean	N	Mean
PCP	Baseline	101,711	0.638	188,480	0.631
	Demo Year 1	93,819	0.629	172,779	0.619
	Demo Year 2	83,745	0.619	151,841	0.616
Specialist	Baseline	57,726	0.392	106,013	0.393
	Demo Year 1	55,735	0.390	101,581	0.390
	Demo Year 2	53,857	0.404	98,218	0.403
Both	Baseline	123,608	0.385	227,194	0.383
	Demo Year 1	113,740	0.381	208,727	0.378
	Demo Year 2	102,892	0.372	186,439	0.372

Adjusted

In adjusted analyses (Exhibit III.6) we see that there are no statistically significant differences in the Bice-Boxerman continuity of care measure between the demonstration and comparison FQHC groups in the first year of the demonstration compared with baseline when we look across all primary care visits (see demonstration*demo Year 1 coefficients), specialist visits, and both. In Year 2 (see demonstration*Year 2 coefficients), we see that the demonstration is associated with a statistically significant but very small (~1 percent) decrease in continuity when looking across all primary care provider visits and when looking at primary care and specialist care together. There is no statistically significant association between the demonstration in Year 2 and continuity for specialist care alone.

Exhibit III.6: Continuity of Care Difference in Differences Regression Results

	PCP Visits Only Estimate (SE)	Specialist Visits Only Estimate (SE)	Both PCP and Specialist Visits Estimate (SE)
Intercept	0.762 (0.008) ***	0.5545 (0.0089) ***	0.5506 (0.0057) ***
Demonstration group	0.014 (0.002) ***	–0.0029 (0.0019)	0.0066 (0.0012) ***
Demo Year 1	–0.007 (0.002)***	–0.0024 (0.0017)	0.0011 (0.0011)
Demo Year 2	–0.005 (0.002)**	0.0019 (0.0019)	–0.0026 (0.0012)
Demonstration*Demo Year 1	–0.001 (0.002)	0.0002 (0.0022)	–0.0021 (0.0014)
Demonstration*Demo Year 2	–0.013 (0.002) ***	–0.0012 (0.0024)	–0.007 (0.0015) ***

***<.0001.
** <.005.
NOTE: Model also controlled for beneficiary, site and area characteristics.

Limitations

Because both PCMHs and FQHCs emphasize a team model of care that may be best measured at the practice level and not the provider level, these results do not provide a comprehensive picture of the potential impact that the demonstration is having on continuity of care. Future analyses will examine whether the demonstration is associated with better continuity of care at the practice level using a version of the Bice-Boxerman index that clusters providers from the same site together with referred providers.

RQ2.2 Do FQHCs Participating in the Demonstration Provide More Timely Delivery of Health Services to Medicare Beneficiaries?

RQ2.2.a Summary: Timely Delivery of Health Care Services (Baseline beneficiary survey)

We used seven CAHPS items to assess timeliness.

- There was no significant difference between demonstration and comparison site users.

RQ2.2.b Analysis: Timely Delivery of Health Care Services (Baseline beneficiary survey)

Seven analyses pertinent to timely delivery of health care services
[Analysis SQ2.1B.1b1] Among all beneficiary survey respondents who had a visit to the attributed clinic in the last 12 months and who *called to get an appointment for routine care*, • 61 percent of demonstration FQHC and 60 percent of comparison FQHC users *always received an appointment as soon as needed* – p=0.5179 for demonstration FQHC vs. comparison FQHC.
[Analysis SQ2.1B.1b1] Among all beneficiary survey respondents who had a visit to the attributed clinic in the last 12 months and who indicated that they *phoned to get an appointment for care right away,* • 55 percent of demonstration FQHC and 55 percent of comparison FQHC users *always received an appointment as soon as needed* – p=0.9675 for demonstration FQHC vs. comparison FQHC.
[Analysis SQ2.1B.1c] Among all beneficiary survey respondents who had a visit to the attributed clinic in the last 12 months and who *called during regular office hours to get a medical question answered*, • 45 percent of demonstration FQHC and 45 percent of comparison FQHC users reported *always getting an answer to their medical question that same day* – p=0.8878 for demonstration FQHC vs. comparison FQHC.
[Analysis SQ2.1B.1d] Among all beneficiary survey respondents who had a visit to the attributed clinic in the last 12 months and who indicated that they *called the office after regular office hours to get a question answered,* • 38 percent of demonstration FQHC and 42 percent of comparison FQHC users reported *always getting an answer to their medical question as soon as needed* – p=0.4665 for demonstration FQHC vs. comparison FQHC.
[Analysis SQ2.1B.1b1] Among all beneficiary survey respondents *who had a visit to the attributed clinic in*

the last 12 months,

- 21 percent of demonstration FQHC and 20 percent of comparison FQHC users reported *always seeing their provider within 15 minutes of their appointment time*
  - p=0.2737 for demonstration FQHC vs. comparison FQHC.

[Analysis SQ2.1A.1c] Among the one-quarter of beneficiary respondents who indicated they *needed after-hours care*,

- 22 percent of demonstration FQHC and 24 percent of comparison FQHC users reported *always getting the care needed during evenings, weekends, or holidays*
  - p=0.3672 for demonstration FQHC vs. comparison FQHC.

[Analysis SQ2.1E.1c] Among respondents who visited their attributed clinic in the last 12 months and who indicated that they *phoned their provider's office to get an appointment for care needed care right away,*

- Slightly more than one-third were given a same-day appointment. 18 percent of demonstration FQHC and 18 percent of comparison FQHC users reported *waiting four or more days for an appointment when needing care right away*
  - p=0.5107 for demonstration FQHC vs. comparison FQHC.

RQ2.3 Do FQHCs Participating in the Demonstration Provide Better or Enhanced Access to Medicare Beneficiaries' PCMH Providers?

RQ2.3.a Summary: Access to Information and Services (Baseline beneficiary survey)

Information-Sharing About Access and Appointment

We assessed how offices share information with beneficiaries about how to access after-hours care.

- Demonstration FQHC users (77%) were not significantly different than comparison FQHC users (74%, p=0.0794).

Access to Information About Services That Occur Between Visits

We measured beneficiary reports of systematic sharing by clinics of reminders about appointments scheduled between office visits (e.g., for tests or treatments).

- Demonstration FQHC users were not significantly different than comparison FQHC users (p=0.1754).

Access to Specialists

Among beneficiaries who visited their attributed clinic within the past 12 months, only slightly more than half reported it was always easy to get an appointment with specialists.

- Demonstration FQHC users were not significantly more likely than comparison FQHC users to report better access to specialists (p=0.2529).

Access to Needed Help with Transportation

We measured beneficiary report of needed help with transportation, and whether help provided by clinics for transportation met their needs.

- Demonstration FQHC users who needed and received help were not significantly more likely than comparison FQHC users to report that help met needs (p=0.0788).

Satisfaction with Provided Interpreter Services

Among beneficiaries reporting that they needed an interpreter in their attributed provider's office, we measured the proportion of beneficiaries who reporting using friends or family members as interpreters when talking with their provider though they preferred not to do so.

- Demonstration FQHC users were significantly more likely than comparison FQHC users to be dissatisfied with using friends or family (p=0.1184).

RQ2.3.b Analysis: Access to Information and Services (Baseline beneficiary survey)

Information-Sharing About Access and Appointment

We observed no difference between demonstration and comparison sites in offices sharing information about how to access after-hours care. However, demonstration FQHC users were significantly more likely than comparison FQHC users to indicate their offices shared information with them about tests, treatments or appointments that needed to occur between office visits.

Information-sharing about how to receive care needed after hours

[Analysis SQ2.1A.1a] Among beneficiary survey respondents who visited an attributed clinic in the past 12 months,

- 76.51 percent of demonstration FQHC and 74.02 percent of comparison FQHC users indicated their provider's office *gave them information about what to do if care was needed during evenings, weekends, or holidays*
  - p=0.0794 for demonstration FQHC vs. comparison FQHC.

Information reminders about tests, treatments or appointments between office visits

[Analysis SQ2.1A.1d] Among beneficiary survey respondents who visited their attributed clinic in the past 12 months,

- 73.18 percent of demonstration FQHC and 71.05 percent of comparison FQHC users indicated that their attributed office reminded them about *tests, treatments, or appointments between office visits*
  - p=0.1754 for demonstration FQHC vs. comparison FQHC.

Access to Specialists

Using a single item that asked beneficiaries who had visited their attributed clinic in the past 12 months how often it was easy to get appointments with specialists, only slightly more than half of FQHC users reported it was always easy to get an appointment with specialists. No significant differences were noted between demonstration FQHC and comparison FQHC users.

Access to appointments with specialists

[Analysis SQ2.1A.1e] Among all beneficiary respondents who visited their attributed clinic in the past 12 months,

- 58.13 percent of demonstration FQHC and 53.66 percent of comparison FQHC users *always reported it was easy to get appointments with specialists*
  - p=0.2529 for demonstration FQHC vs. comparison FQHC.

Access to Needed Help with Transportation

Among the 10–12 percent of respondents who indicated they needed help with transportation, there were no significant differences between demonstration and comparison FQHC users in respondents' report that their provider's office helped with transportation. Among those who noted that their provider's office helped with transportation, more than half reported that the help met their needs with no significant differences between demonstration and comparison site FQHC users.

Access to needed help with transportation

[Analysis SQ2.1a.2a] Among beneficiary respondents who visited their attributed clinic in the last 12 months,

- 12.00 percent of demonstration FQHC and 11.49 percent of comparison FQHC users indicated they *needed help with transportation to visits to their provider's office*.
  - p=0.7959 for demonstration FQHC vs. comparison FQHC.

[Analysis SQ2.1a.2b] Among those who indicated that they needed help with transportation,

- 37.97 percent of demonstration FQHC users and 29.33 percent of comparison FQHC users indicated their *provider's office helped with transportation*
  - p=0.2813 for demonstration FQHC vs. comparison FQHC.

[Analysis SQ2.1a.2c] While most (62.03 percent of demonstration FQHC users and 70.67 percent of comparison FQHC users indicated their provider's office did not help with transportation, conditional on respondents indicating they received help with transportation,

- 70.74 percent of demonstration FQHC users and 53.28 percent of comparison FQHC users noted the provided transportation help met their needs.
  - p=0.0788 for demonstration FQHC vs. comparison FQHC.

As a supplement to this beneficiary survey analysis, the CASE survey shows that 12 percent of demonstration FQHC users indicated needing help with transportation, but only slightly more than one-third noted their provider's office helping with

transportation. In this context, it is useful to know that almost 30 percent of CASE respondents reported that their sites were making efforts to increase the availability of transportation to and from sites.

Access to Needed Interpreter

There was no significant difference between demonstration FQHC users and comparison FQHC users with respect to whether they always used an interpreter provided by the office to help talk with the provider.

Use of friends or family members as interpreters among beneficiaries who preferred a different type of interpreter

[Analysis SQ2.1A.3e] Among beneficiaries who used a friend or family member,

- 35.47 percent for demonstration FQHC and 25.55 percent for comparison FQHC users *preferred not to use friends or family members as interpreters*
  - p=0.1184 for demonstration FQHC vs. comparison FQHC.

RQ2.3.c Summary: Access to Information and Services (Baseline CASE survey)

As a supplement to beneficiary survey reports about transportation, CASE analyses focus on transportation and other aspects of services delivered outside office visits. While the effects of demonstration participation on access cannot be estimated until the conclusion of the demonstration, cross-sectional relationships between RAS levels and these dimensions of access may foreshadow demonstration effects.

Nearly 30 percent of respondents reported that their sites were making efforts to increase the availability of transportation to and from the site, nearly two-thirds percent reported that their sites were making efforts to increase the number of office visits with patients , nearly 25 percent reported that their sites were making efforts to increase the amount of patient care via telephone , and nearly one-third reported that their sites were making efforts to increase the amount of patient care via email . None of these dimensions of access were statistically significantly associated with RAS levels measured immediately prior to the fielding of baseline CASE surveys.

To date, no differences in site reports of improved access are associated with RAS scores, but the emphasis by some sites on improving a variety of measures of access is encouraging. Our team in the coming year will be analyzing changes in CASE reports of efforts to improve access as a predictor of improved RAS scores and NCQA Level 3 recognition.

RQ2.4 Do FQHCs Participating in the Demonstration Provide Improved Adherence to Evidence-Based Guidelines?

RQ2.4.a Summary: Adherence to Evidence-Based Guidelines (Baseline beneficiary survey)

We assessed use of evidence-based guidelines using six domains of explicit process measures and found no significant differences between demonstration and comparison FQHC users for four of the six domains. However, for the remaining two (counseling for healthy eating and exercise, and mental health assessments), beneficiary reports indicate greater adherence to evidence-based recommendations by demonstration FQHCs than by comparison FQHC sites.

RQ2.4.b Analysis: Adherence to Evidence-Based Guidelines (Baseline beneficiary survey)

Immunization

We assessed vaccination use among 65- to 85-year-old beneficiaries for influenza, pneumococcal pneumonia, and shingles.

- Demonstration FQHC users were not different from comparison FQHC users for influenza vaccines (p=0.3532), pneumonia (p=0.6033), or shingles vaccinations (p=0.2628).

Use of influenza, pneumonia, and shingles vaccinations

[Analysis SQ2.1C.1a] Among the subset of beneficiary respondents ages 65–85 years,

- 67.61 percent of demonstration FQHC and 64.19 percent of comparison FQHC users reported *receipt of the influenza vaccine during the year prior to the baseline beneficiary survey*
  - p=0.3532 for demonstration FQHC vs. comparison FQHC.

[Analysis SQ2.1C.1b] Among beneficiary respondents ages 65–85 years,

- 71.17 percent of demonstration FQHC and 69.28 percent of comparison FQHC users reported *having ever received a pneumonia shot (the pneumococcal vaccine)*
  - p=0.6033 for demonstration FQHC vs. comparison FQHC.

[Analysis SQ2.1C.1c] Among beneficiary respondents ages 65–85 years,

- 18.50 percent of demonstration FQHC and 22.06 percent of comparison FQHC users reported *receipt of a shingles (herpes zoster) vaccine*;
  - p=0.2628 for demonstration FQHC vs. comparison FQHC.

Primary Prevention of Cardiovascular Disease with Aspirin Use

Among beneficiaries with cardiovascular disease or diabetes, we found no significant differences between demonstration and comparison group users regarding use of, or counseling about use of, aspirin for primary or secondary prevention of cardiovascular disease and stroke.

- Demonstration FQHC users were not different from comparison FQHC users when cardiovascular disease or diabetes was defined by claims (p=0.9930) or by beneficiary survey (p=0.4704).

Colorectal Cancer Screening

We defined guideline-concordant receipt of colorectal cancer screening for adults ages 55–85 years as any of the following: (1) colonoscopy within 10 years, (2) sigmoidoscopy within five years, or (3) fecal occult blood test within two years.

Colorectal cancer screening among beneficiaries aged 55–75 years

[SQ2.1C.5] Among all beneficiary respondents aged 55–75 years,

- 64.88 percent of demonstration FQHC and 61.64 percent of comparison FQHC users reported *guideline-concordant colorectal screening*
  - Demonstration FQHC users were not different from comparison FQHC users (p=0.3950).

Engaging Beneficiaries Who Smoke in Discussions About Three Evidence-Based Smoking Cessation Interventions

The Centers for Disease Control recommends clinicians regularly advise smoking patients to quit, recommend medications to assist the patient in quitting, and discuss or provide methods and strategies beyond medication to assist the patient with quitting smoking or using tobacco. Among smoking respondents who visited their attributed

provider within the last 12 months, 17 percent of demonstration FQHC and 14 percent of comparison FQHC users report receiving none of the three recommended interventions, while 46 percent of demonstration FQHC and 43 percent of comparison FQHC users report receiving all three of these interventions.

- Demonstration FQHC users were not different from comparison FQHC users (p=0.7568).

Counseling Regarding Weight Loss, Exercise, and Eating Right

Counseling adults regarding weight loss, exercising regularly, and eating right has been recommended to improve cardiovascular and overall health.

- Demonstration FQHC users were not different from comparison FQHC users (p=0.1794).

Mental Health Interventions

With the increasing recognition that mental health is an important predictor of physical illness, poor health status, adverse outcomes, and health care utilization, we used three process items to assess whether survey respondents recognized their provider's office as assessing their mood or other mental or emotional health concerns. These include provider evaluation during the prior 12 months of beneficiary mood, stress levels, and precipitants of stress.

Clinician evaluation of beneficiary's mood, levels of stress, and precipitants of stress

[SQ2.1C.6] Among all beneficiary respondents who visited their attributed clinic in the last 12 months,

- 38 percent of demonstration FQHC and 43 percent of comparison FQHC users reported *never talking with their providers about any of these mental health topics*, while 29 percent of demonstration FQHC and 24 percent of comparison FQHC users reported *receiving all three of these interventions.*
  - Demonstration FQHC users were significantly more likely than comparison FQHC users to receive these mental health assessments (**p=0.0039).**

RQ2.4.c Summary: Adherence to Evidence-Based Guidelines (Baseline CASE Survey)

The baseline CASE survey provides data pertinent to one of our evidence-based domain measures, the provision of nutrition and weight loss services. Providing such services is concordant with guidelines for several patient groups salient to the demonstration (e.g., those with obesity, diabetes, and osteoarthritis of weight-bearing joints). Baseline CASE survey analyses show that approximately 50 percent of respondents reported that their sites were making efforts to increase the availability of

nutrition or weight loss services. Thus far, these reported rates of efforts to change have not varied by RAS score.

RQ2.4.d Summary: Adherence to Evidence-Based Guidelines (Claims data)

We assessed the impact of the demonstration on four process measures for Medicare beneficiaries with diabetes and one process measure for Medicare beneficiaries with ischemic vascular disease. For diabetic patients, we assessed the percentage receiving HbA1c tests, the percentage receiving retinal eye exams, the percentage receiving nephropathy testing, and the percentage receiving LDL cholesterol testing in the past year. For patients with ischemic vascular disease, we assessed the percentage receiving a lipid test over the past year. For this analysis, the FQHC demonstration cohort was compared to the FQHC comparison cohort for the ITT cohort.

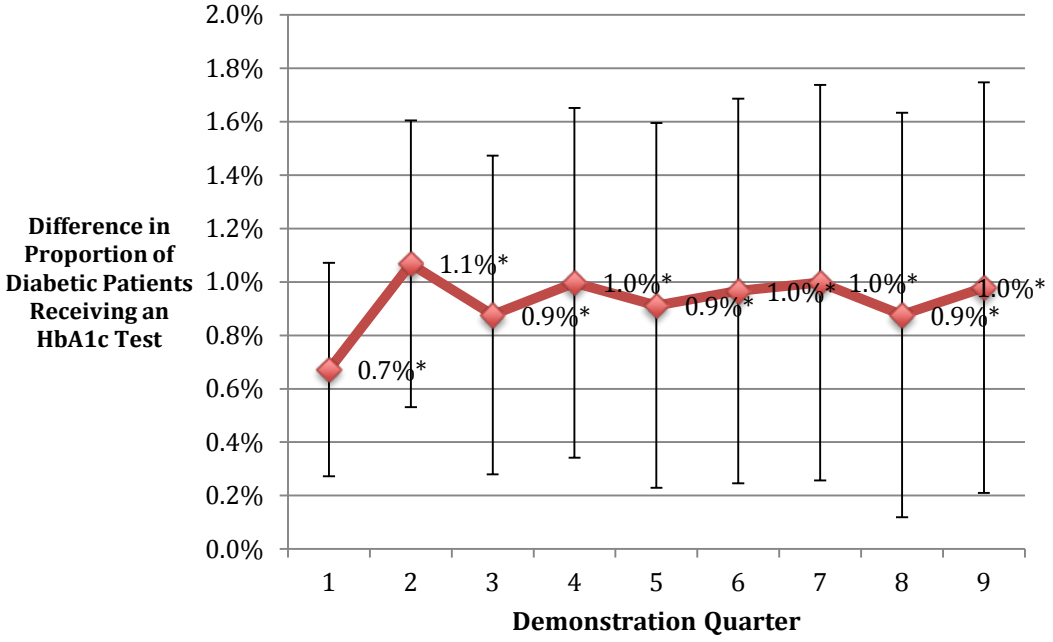
The results for diabetic patients indicate that the FQHC demonstration group generally did the same or better than the FQHC comparison group in terms of recommended diabetic testing. The results for ischemic vascular disease indicate that the FQHC demonstration group performed similarly to the comparison group in terms of adherence to recommended lipid testing. All differences during the demonstration are relative to differences in the baseline period.

RQ2.4.e Analysis: Adherence to Evidence-Based Guidelines (Claims data)

HbA1c testing for diabetic patients

Exhibit III.7 shows the impact of the demonstration on the percentage of Medicare beneficiaries with diabetes receiving an HbA1c test in the past year when the FQHC demonstration cohort was compared to the FQHC comparison cohort in ITT analyses.. Regression results indicate that in demonstration quarters 1–9, the difference between the demonstration sites and the comparison sites was significantly greater than the difference observed between these two groups in the baseline period. Over the nine demonstration quarters, demonstration FQHCs provided HbA1c tests to an additional 1 percent of diabetic beneficiaries relative to comparison FQHCs for each quarter, adjusting for differences in the baseline period.

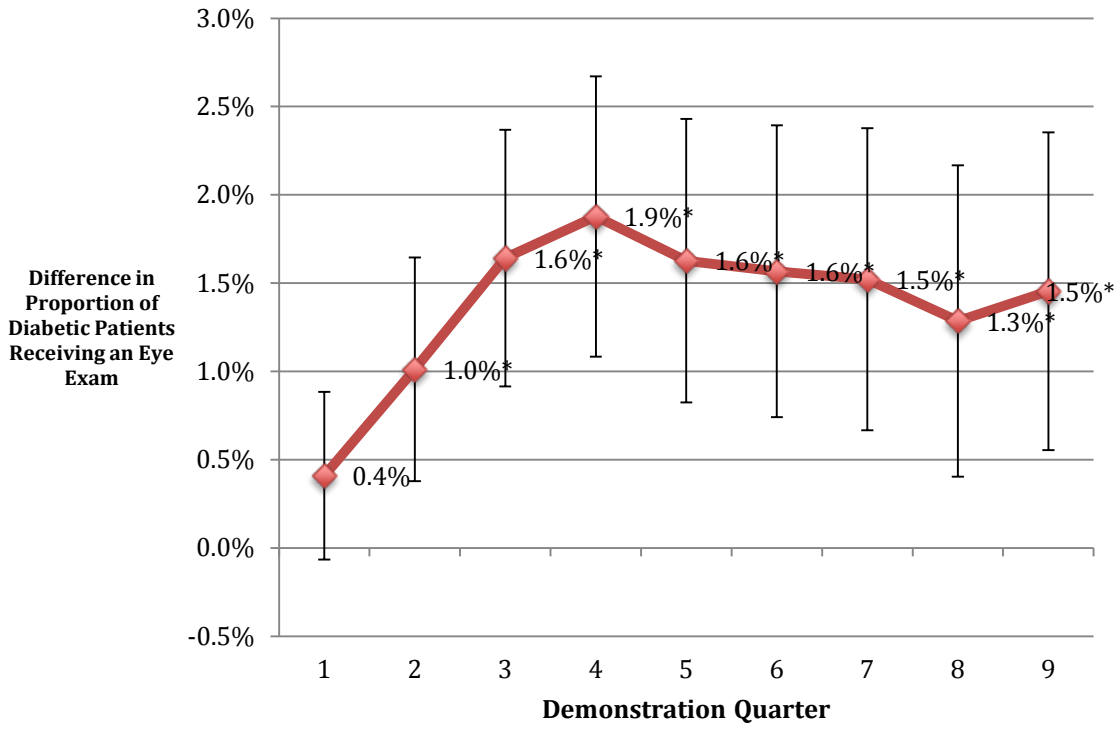
Exhibit III.7: Difference in Process Measures per Demonstration Quarter (Demonstration FQHCs vs. Comparison FQHCs) Based on Regression Analysis: HbA1c Testing for Diabetic Patients



Eye Examination for Diabetic Patients

Exhibit III.8 shows the impact of the demonstration on the percentage of Medicare beneficiaries with diabetes who received retinal eye exams in the past year when the FQHC demonstration cohort was compared to the FQHC comparison cohort in ITT analyses. Regression results indicate that FQHC demonstration sites had rates of eye exams that were approximately 1 to 2 percentage points higher in quarters 2 to 9 than FQHC comparison sites.

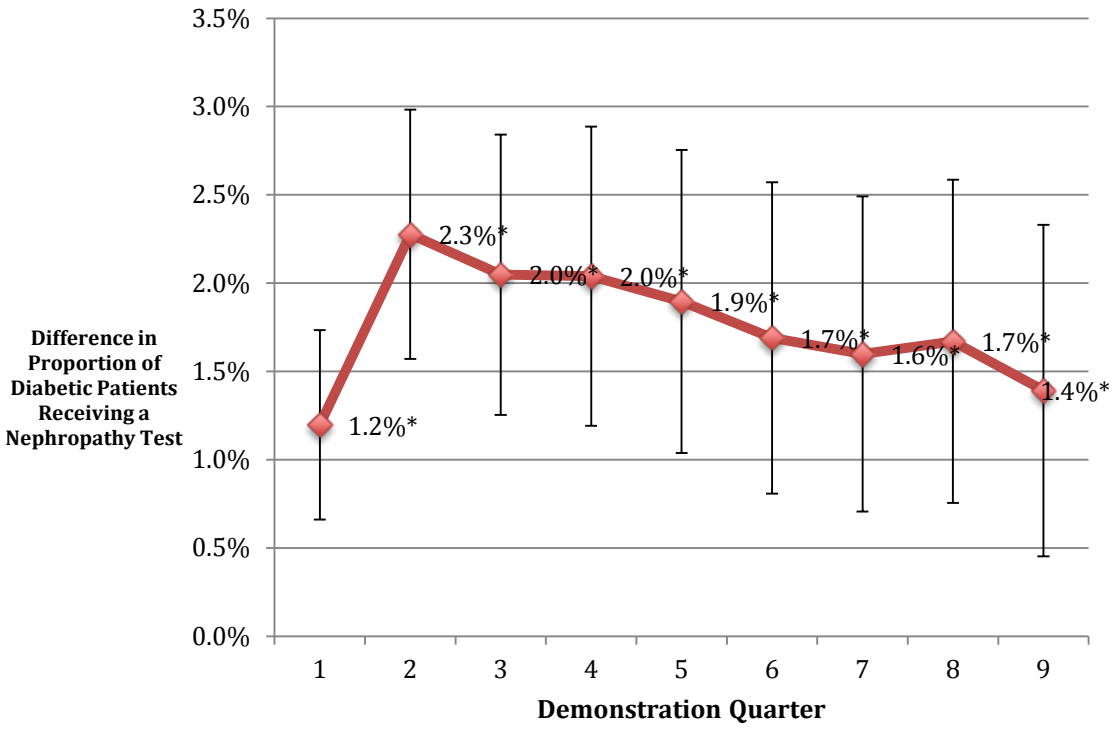
Exhibit III.8: Difference in Process Measures per Demonstration Quarter (Demonstration FQHCs vs. Comparison FQHCs) Based on Regression Analysis: Eye Exam for Diabetic Patients



Nephropathy Testing for Diabetic Patients

Exhibit III.9 shows the impact of the demonstration on the percentage of diabetics receiving nephropathy testing in the past year for ITT analyses. Regression results indicate that in each of the nine demonstration quarters, the FQHC demonstration sites had higher rates of nephropathy testing by 1–2 percentage points more than the FQHC comparison sites, relative to the difference observed between the two groups in the baseline period.

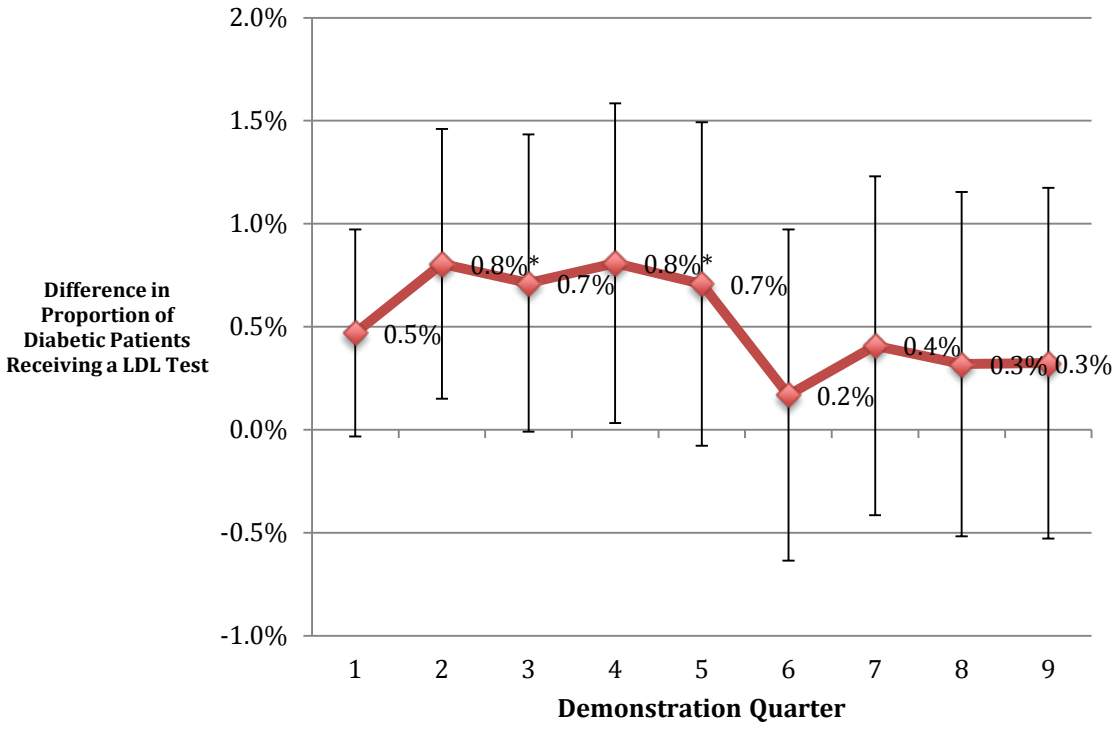
Exhibit III.9: Difference in Process Measures per Demonstration Quarter (Demonstration FQHCs vs. Comparison FQHCs) Based on Regression Analysis: Nephropathy Test for Diabetic Patients



LDL Testing for Diabetic Patients

Exhibit III.10 shows the impact of the demonstration on the percentage of Medicare beneficiaries with diabetes receiving LDL cholesterol testing in the past year when the FQHC demonstration cohort was compared to the FQHC comparison cohort in primary ITT analyses. Regression results indicate that differences between the groups were statistically significant only in quarters 2 and 4. In these quarters, diabetic beneficiaries attributed to demonstration FQHCs had rates of LDL cholesterol testing that were 0.8 percentage points higher than beneficiaries attributed to comparison FQHCs, relative to differences at baseline.

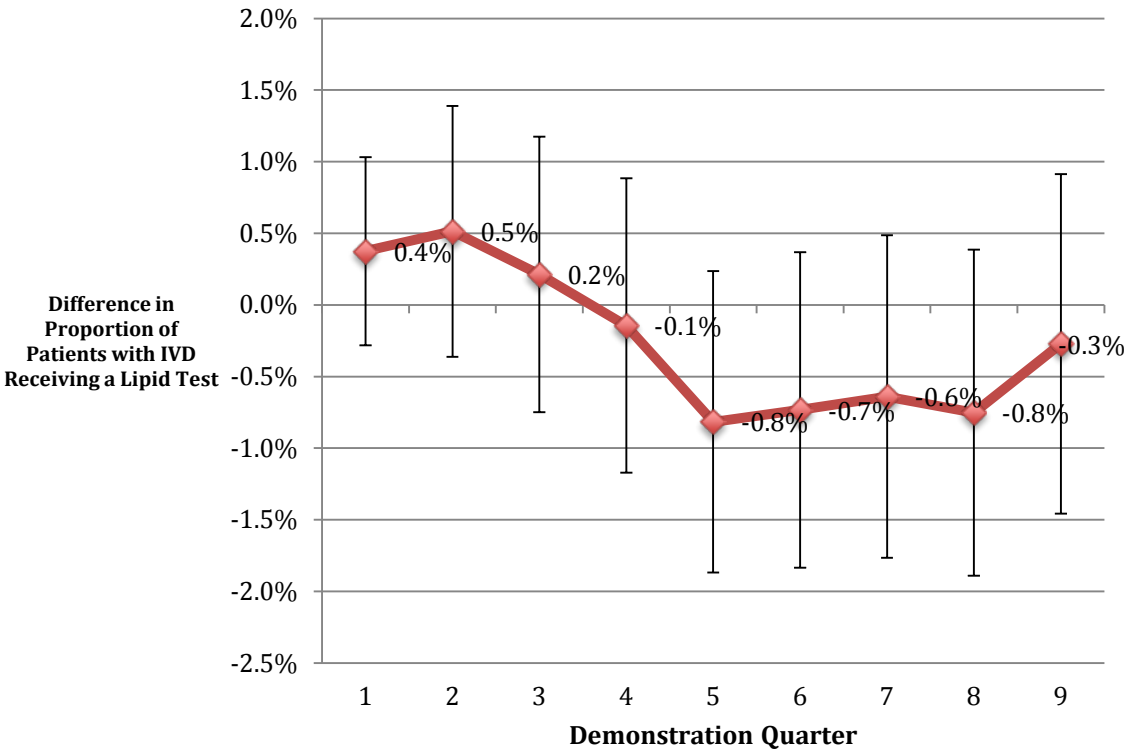
Exhibit III.10: Difference in Process Measures per Demonstration Quarter (Demonstration FQHCs vs. Comparison FQHCs) Based on Regression Analysis: LDL Test for Diabetic Patients



Lipid Profile Testing for Patients with Ischemic Vascular Disease

Exhibit III.11 shows the impact of the demonstration on the percentage of beneficiaries with ischemic vascular disease receiving a blood lipid profile in the past year when the FQHC demonstration cohort was compared with the FQHC comparison cohort in ITT analyses. Regression results indicate that there were no differences between the groups in demonstration quarters 1–9 that were statistically different from the difference observed between the groups in the baseline period (all p-values >0.05).

Exhibit III.11: Difference in Process Measures per Demonstration Quarter (Demonstration FQHCs vs. Comparison FQHCs) Based on Regression Analysis: Lipid Test for Ischemic Vascular Disease (IVD) Patients



RQ2.5 Do FQHCs Participating in the Demonstration Provide Better Quality of Care to Medicare Beneficiaries?

RQ2.5.a Summary: Quality of Care Rating (Baseline beneficiary survey)

We used three metrics of beneficiary-rated quality, to assess care received by Medicare beneficiaries: (1) ratings of their attributed providers, (2) ratings of specialty providers, and (3) ratings of "helpful, courteous, and respectful staff."

Across all three metrics:

- Demonstration FQHC users were not different from comparison FQHC users ($p>0.05$).

RQ2.5.b Analysis: Quality of Care Rating (Baseline beneficiary survey)

Beneficiary ratings of their attributed providers

[Analysis SQ2.1D.1a] Among beneficiary respondents who visited their attributed clinic within the last 12 months, using a 10-point CAHPS rating scale where 0 is the worst provider possible and 10 is the best,

- Respondents assigned a mean (SD) score of 8.65 (3.94) for demonstration FQHCs and 8.59 (3.61) for comparison FQHCs (for the mean rating)
  - p=0.4027 for demonstration FQHC vs. comparison FQHC.

Beneficiary ratings of specialty providers

[Analysis SQ2.1D.1b] Among beneficiary respondents who visited their attributed clinic within the last 12 months and saw a specialist, using a 10-point CAHPS rating scale where 0 is the worst specialist possible and 10 is the best,

- Respondents assigned a mean (SD) score of 8.91 (3.26) for demonstration FQHCs, and 8.78 (2.79) for comparison FQHCs
  - p=0.3819 for demonstration FQHC vs. comparison FQHC.

Beneficiary Ratings of 'Helpful, Courteous, and Respectful Office Staff'

We used the Clinician and Group Consumer Assessment of Healthcare Providers and Systems (CG-CAHPS) "Helpful, Courteous, and Respectful Office Staff" two-item and aggregated scale to assess this domain. The two items composing this scale are:

1. In the last 12 months, how often were clerks and receptionists at this provider's office as helpful as you thought they should be?

2. In the last 12 months, how often did clerks and receptionists at this provider's office treat you with courtesy and respect?

Beneficiary ratings of "helpful, courteous, and respectful office staff"

[Analysis SQ2.1I.1. Summary] Among beneficiary respondents who visited their attributed clinic within the last 12 months, using a higher value to represent more frequent exposure to valued communications,

- Respondents assigned a mean (SD) score of 89.8 (31.5) for demonstration FQHCs and 89.0 (28.8) for comparison FQHCs
  - p=0.1972 for demonstration FQHC vs. comparison FQHC.

RQ2.6 Are Medicare Beneficiaries, Their Family Members, and/or Their Caregivers Served by Participating FQHCs Able to Participate More Effectively in Decisions Concerning Their Care?

RQ2.6.a Summary: Effective Participation in Decisions (Baseline beneficiary survey)

We assessed support provided to Medicare beneficiaries for effectively participating in decisions concerning their care by focusing on beneficiaries' report of communications with providers about use of prescription medications. We analyzed beneficiary report about provider discussions regarding starting or stopping a prescription medicine, reasons the beneficiary might want to take a medicine, reasons the beneficiary might not want to take a medicine, and what decision the beneficiary thought best for themselves. Using these four measures of effective participation in health decisions,

- Demonstration FQHC users were not different from comparison FQHC users ($p>0.05$ for all four measures).

RQ2.7 Are Medicare Beneficiaries Served by the Participating FQHCs Better Able to Self-Manage Their Health Conditions or More Likely to Engage in Healthy Behaviors?

Using the beneficiary survey data, we address the question of self-management using scales pertinent to both self-management and to how providers support patients in taking care of their own health.

RQ2.7.a Summary: Provider-Beneficiary Communication About Self-Management (Baseline beneficiary survey)

We assessed more details about how providers instructed beneficiaries about their specific illness or condition using the four-item CAHPS Health Literacy Disease Self-Management scale. The scale asks about "how often" each of the following characteristics of communications between the beneficiary and provider occurred during the last 12 months: (1) instructions were easy to understand, (2) the provider asked you to describe how you were going to follow these instructions, (3) the provider asked whether you would have any problems doing what you need to do to take care of this illness or health condition, and (4) the provider explained what to do if this illness or health condition got worse or came back.

- Demonstration FQHC users were not different from comparison FQHC users ($p=0.9550$).

RQ2.7.b Analysis: Educating Beneficiaries for Self-Management (Baseline beneficiary survey)

Provider-beneficiary communication about self-management of health conditions using CAHPS Health Literacy Disease Self-Management *scale*

[Analysis SQ2.2H.1b Summary] Among beneficiary respondents who visited their attributed provider for a specific illness or for any health condition within the last 12 months,

- The mean communication about self-management score was 79.5 (38.7) for demonstration FQHC and 79.3 (33.4) for comparison FQHC users
  - p=0.9550 for demonstration FQHC vs. comparison FQHC.

RQ2.7.c Summary: Provider Support for Beneficiaries Caring for Their Own Health[24]

We also assessed support provided to Medicare beneficiaries for taking care of their own health using the CAHPS PCMH Items Set, "Providers Support You in Taking Care of Your Own Health." This scale includes two pertinent dichotomous items:

1. In the last 12 months, did anyone in this provider's office talk with you about specific goals for your health?
2. In the last 12 months, did anyone in this provider's office ask you if there are things that make it hard for you to take care of your health?

Using the overall scale score to measure provider support for beneficiaries caring for their own health:

- Demonstration FQHC users were significantly more likely than comparison FQHC users to indicate their providers supported them to care for their own health (**p=0.0110**).

RQ2.7.d Analysis: Provider Support for Beneficiaries Caring for their Own Health (Baseline beneficiary survey)

Provider supports for beneficiaries taking care of their own health using CAHPS scale

[Analysis SQ2.2H.1a Summary] Across these two items, among beneficiary respondents who visited their attributed provider within the last 12 months,

- We note a mean (SD) CAHPS PCMH "Provider Supports You in Taking Care of Your Own Health" scale rating of 50.68 (78.3) for demonstration FQHC and 47.45 (76.6) for comparison FQHC users
  - **p=0.0110** for demonstration FQHC vs. comparison FQHC.

The higher mean value for demonstration FQHCs is explained by their providers being more likely to query beneficiaries about their goals; no differences were noted in

[24] CAHPS. (2011). *About the patient-centered medical home item set.* Document No. 1314. Accessed December 4, 2014: https://cahps.ahrq.gov/surveys-guidance/survey4.0-docs/1314_About_PCMH.pdf

the prevalence of providers asking beneficiaries about things that make it hard for them to care for their health.

RQ2.7.e Summary: Provider Follow-Up on Test Results (Baseline beneficiary survey)

Using the single CAHPS item pertinent to communications about test results:

- Demonstration FQHC users were not different from comparison FQHC users (p=0.8445).

RQ2.7.f Analysis: Provider Follow-Up on Test Results (Baseline beneficiary survey)

Provider's office communication with beneficiary about test results
[Analysis SQ2.1H.1g] Among beneficiary respondents who visited their attributed provider within the last 12 months and who had a provider order a test within the last 12 months, • 8.33 percent of demonstration FQHC and 8.93 percent of comparison FQHC users reported *nobody from their provider's office ever followed up to give them those results*, while 66.06 percent of demonstration FQHC and 64.94 percent of comparison FQHC users reported *somebody from their provider's office always followed up to give them those results* – p=0.8445 for demonstration FQHC vs. comparison FQHC.

Comparison data from the 2013 CG-CAHPS database show an average of 70 percent of respondents reporting somebody from their provider's office always followed up to give them those results.[25]

RQ2.8 Do FQHCs Participating in the Demonstration Provide Better Experiences with the Health Care System for Medicare Beneficiaries and Their Families and Caregivers?

RQ2.8.a Summary: Beneficiary Experiences (Baseline beneficiary survey)

We used the six-item CG-CAHPS "How Well Providers Communicate with Patients" scale to assess beneficiary experiences with health care system quality to assess care received by Medicare beneficiaries. The six items this scale comprises are:

1. In the last 12 months, how often did this provider explain things in a way that was easy to understand?
2. In the last 12 months, how often did this provider listen carefully to you?
3. In the last 12 months, how often did this provider give you easy-to-understand information about these health questions or concerns?

[25] CAHPS. (2014). *2013 chartbook: What patients say about their health care providers and medical practices*. Rockville, Md.: Westat, p. 2, Table 1-1. Accessed December 2, 2014: https://www.cahps.ahrq.gov/cahps-database/about/2013cahpscliniciangroupchartbook.pdf

4. In the last 12 months, how often did this provider seem to know the important information about your medical history?
5. In the last 12 months, how often did this provider show respect for what you had to say?
6. In the last 12 months, how often did this provider spend enough time with you?

Using the overall CG-CAHPS scale score to measure beneficiary reports of their experiences with the health care system:

- Demonstration FQHC users were not different from comparison FQHC users for the overall scale (p=0.2290).

RQ2.8.b Analysis: Beneficiary Experiences (Baseline beneficiary survey)

Across six measures of provider communications with patients, beneficiaries were most likely to report their providers always showed respect for what they had to say (80.30 percent of demonstration FQHC, 78.62 percent of comparison FQHC users). Respondents were least likely to report their providers always seemed to know the important information about their medical history (65.84 percent of demonstration FQHC, 64.55 percent of comparison FQHC).

Beneficiary experiences with the health care system measured with CAHPS How Well Providers Communicate with Patients *scale*
[Analysis SQ2.1I.1a] Among beneficiary survey respondents who visited their attributed clinic within the last 12 months, • Fewer than 2.5 percent of respondents across all three cohorts indicated *providers never explained things in a way that was easy to understand*, while 73.21 percent of demonstration FQHC and 71.66 percent of comparison FQHC users reported providers always *explained things in a way that was easy to understand* – p=0.2215 for demonstration FQHC vs. comparison FQHC.
[Analysis SQ2.1I.1b] Among beneficiary survey respondents who visited their attributed clinic within the last 12 months, • Fewer than 2.5 percent of respondents across all cohorts indicated *providers never listened carefully to them*, while 76.12 percent of demonstration FQHC and 75.23 percent of comparison FQHC users reported *providers always listened carefully to them* – p=0.7450 for demonstration FQHC vs. comparison FQHC.
[Analysis SQ2.1I.1d] Among beneficiary survey respondents who visited their attributed clinic within the last 12 months and who discussed health questions or concerns, • Fewer than 2.5 percent of respondents indicated *providers never gave them easy-to-understand information about their health questions or concerns,* while 70.97 percent of demonstration FQHC and 68.95 percent of comparison FQHC users reported *providers always explained things in a way that was easy to understand* – p=0.4538 for demonstration FQHC vs. comparison FQHC.
[Analysis SQ2.1I.1e] Among beneficiary respondents who visited their attributed clinic in the last 12 months, • 3.49 percent of demonstration FQHC and 3.66 percent of comparison FQHC users reported that *their provider*

never seemed to know the important information about their medical history, while 65.84 percent of demonstration FQHC and 64.55 percent of comparison FQHC users reported that *their provider always seemed to know the important information about their medical history*
  - p=0.8256 for demonstration FQHC vs. comparison FQHC.

[Analysis SQ2.1I.1f] Among beneficiary survey respondents who visited their attributed clinic in the last 12 months,

- Fewer than 2.75 percent of respondents across all three cohorts indicated *providers never showed respect for what they had to say*, while 80.30 percent of demonstration FQHC and 78.62 percent of comparison FQHC users reported *providers always showed respect for what they had to say*
  - p=0.5158 for demonstration FQHC vs. comparison FQHC.

[Analysis SQ2.1I.1g] Among beneficiary survey respondents who visited their attributed clinic within the last 12 months,

- Fewer than 3.60 percent of respondents across all three cohorts indicated *providers never spend enough time with them*, while 66.97 percent of demonstration FQHC and 66.19 percent of comparison FQHC users reported providers *always spend enough time with you*
  - p=0.5654 for demonstration FQHC vs. comparison FQHC.

[Analysis SQ2.1I.1 Summary] Among beneficiary respondents who visited their attributed clinic within the last 12 months,

- Using a higher value to represent more frequent valued communications, respondents assigned a mean (SD) score of 90 (28.6) for demonstration FQHCs and 89.3 (27.5) for comparison FQHCs.
  - p=0.2290 for demonstration FQHC vs. comparison FQHC

RQ2.9 How Does FQHC Participation in the Demonstration Affect Utilization of Services Covered by Medicare?

RQ2.9.a Summary: Utilization of Services (Baseline beneficiary survey)

We assessed beneficiary utilization during the last 12 months for nine services. For six of the listed services, demonstration FQHC utilization was not different from utilization at FQHC comparison sites. However, utilization of specialist visits and mental health services were significantly higher for demonstration FQHCs. Utilization of dental services was significantly lower in demonstration FQHCs than in the comparison group. The utilization services studied include:

1. provider visits during last 12 months
   - Demonstration FQHC users (92 percent) were no different from comparison FQHC users (93 percent) with respect to at least one visit during the last 12 months, (p=0.4879) and no different with respect to the distribution of number of visits (p=0.4672).
2. hospitalization
   - Demonstration FQHC users (28 percent) were no different from comparison FQHC user (30 percent, p=0.3344).
3. vision services for adults

- Demonstration FQHC users (53 percent) were no different from comparison FQHC users (58 percent, p=0.0800) and among diabetic beneficiaries, demonstration FQHC users (65.38 percent) were no different from comparison FQHC users (70.82 percent, p=0.1165).

4. services for hearing assessment
   - Demonstration FQHC users (10 percent) were not different from comparison FQHC users (12 percent, p=0.4021).
5. provider-ordered blood tests, x-rays, or other tests during last 12 months
   - Demonstration FQHC users (89 percent) were not different from comparison FQHC users (88 percent, p=0.3328).
6. use of prescription medications
   - Demonstration FQHC users (96 percent) were not different from comparison FQHC users (96 percent p=0.4533).
7. use of specialist visits

   Demonstration FQHC users (65 percent) were significantly more likely to use specialist visits than comparison FQHC users (61 percent, **p=0.0298**).
8. use of mental health specialty services at least once within the last 12 months
   - Demonstration FQHC users (20 percent) were significantly more likely to use mental health specialty services than comparison FQHC users (14 percent, **p=0.0439**).
9. use of dental services at least once within the last 12 months
   - Demonstration FQHC users (37 percent) were significantly less likely to use dental services than comparison FQHC users (39 percent, **p=0.0126**).

RQ2.9.b Summary: Utilization of Services (Claims data)

We assessed the impact of the demonstration on the number of acute care inpatient admissions, number of inpatient admissions for chronic ACSCs, and the number of ED visits, per 1,000 beneficiaries, as well as the percentage of inpatient admissions that result in an unplanned readmission within 30 days, for each quarter of the demonstration period.

Across nine quarters comparing demonstration FQHCs with comparison FQHCs adjusted for differences at baseline, we noted significantly more utilization for three of four claims-based utilization measures. Specifically, there was no demonstration quarter in which demonstration site users had significantly less utilization than comparison sites. In contrast with comparison FQHC users, demonstration FQHC users had significantly more use of:

- hospital admissions in two of nine demonstration quarters

- admissions for ACSCs in none of nine demonstration quarters
- unplanned hospital readmissions in one of nine demonstration quarters, and
- emergency room visits in five of nine demonstration quarters.

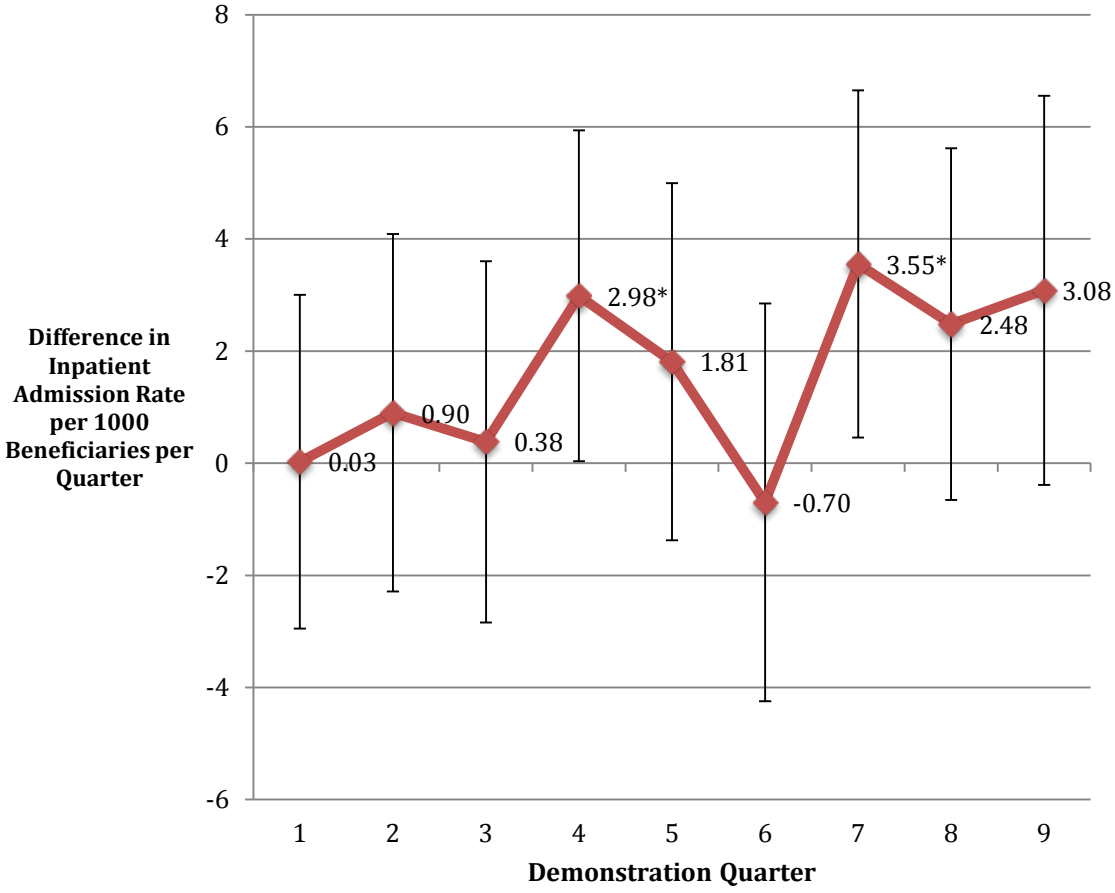
RQ2.9.c Analysis: Utilization of Services (Claims data)

Inpatient Admissions[26]

Exhibit III.12 shows the impact of the demonstration on the number of acute care inpatient admissions per 1,000 beneficiaries per quarter of the demonstration period when the FQHC demonstration cohort was compared to the FQHC comparison cohort in ITT analyses. Regression results indicate that in demonstration quarters 4 and 7, there were 3.0 and 3.6 additional admissions per 1,000 beneficiaries per quarter in the demonstration group versus the comparison group (p-values <0.05). There were no differences between the demonstration and comparison groups in any other demonstration quarters that were statistically different from the differences observed in the baseline period (all p-values >0.05).

[26] For inpatient admissions and all other measures of utilization listed in this section, the outcomes were assessed for our ITT sample in which the FQHC demonstration cohort was compared to the FQHC comparison cohort without late entrants and sample. Future reports will show analyses in which late entrants were added to the FQHC ITT sample.

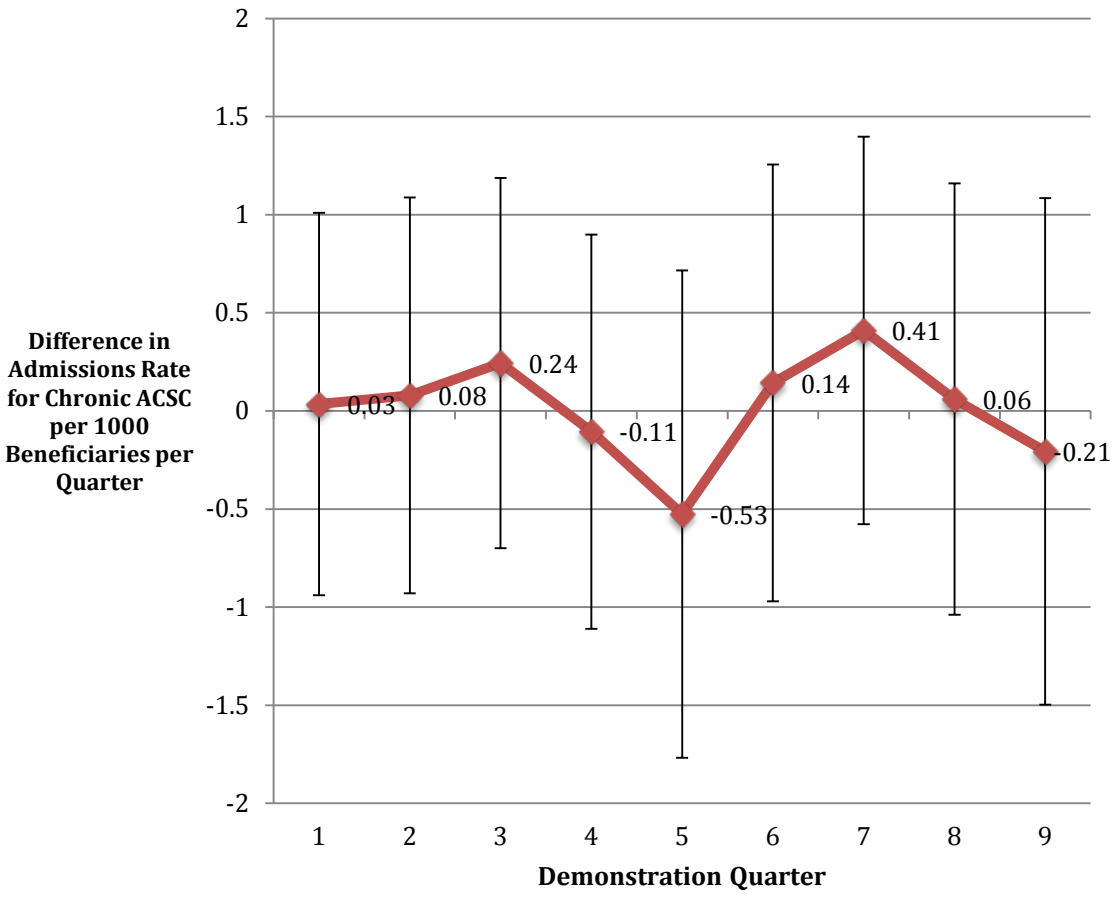
Exhibit III.12: Difference in the Number of Inpatient Admissions per 1,000 Beneficiaries per Demonstration Quarter (Demonstration FQHCs vs. Comparison FQHCs) Based on Regression Analysis (ITT Cohort Only)



Inpatient Admissions for Chronic ACSCs

Exhibit III.13 shows the impact of the demonstration on the number of inpatient admissions for chronic ACSCs per 1,000 beneficiaries per quarter of the demonstration period when the FQHC demonstration cohort was compared to the FQHC comparison cohort in ITT analyses. Regression results indicate that there were no differences between the groups in demonstration quarters 1–9 that were statistically different from the differences observed in the baseline period (all p-values >0.05).

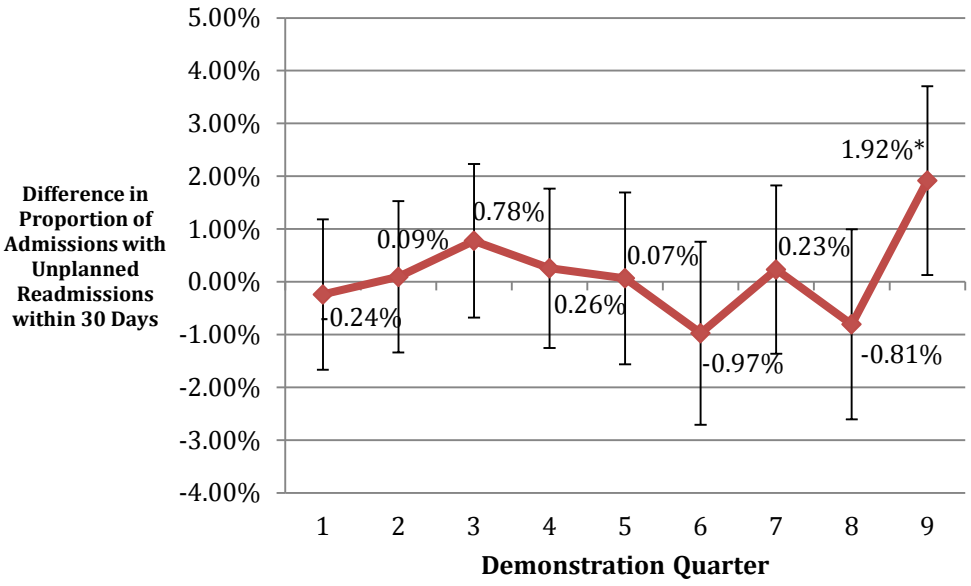
Exhibit III.13: Difference in the Number of Inpatient Admissions for Chronic ACSCs per 1,000 Beneficiaries per Demonstration Quarter (Demonstration FQHCs vs. Comparison FQHCs) Based on Regression Analysis (ITT Cohort Only)



Unplanned Hospital Readmissions

Exhibit III.14 shows the impact of the demonstration on the percentage of inpatient admissions that result in an unplanned hospital readmission within 30 days when the FQHC demonstration cohort was compared to the FQHC comparison cohort in ITT analyses. Regression results indicate that there were no differences between the groups in demonstration quarters 1–8 that were statistically different from the difference observed between the groups in the baseline period (all p-values >0.05). In Quarter 9, when compared with the FQHC comparison group, there was a statistically significant 1.92-percent increase in FQHC beneficiaries from the demonstration group that were readmitted within 30 days, adjusted for differences at baseline.

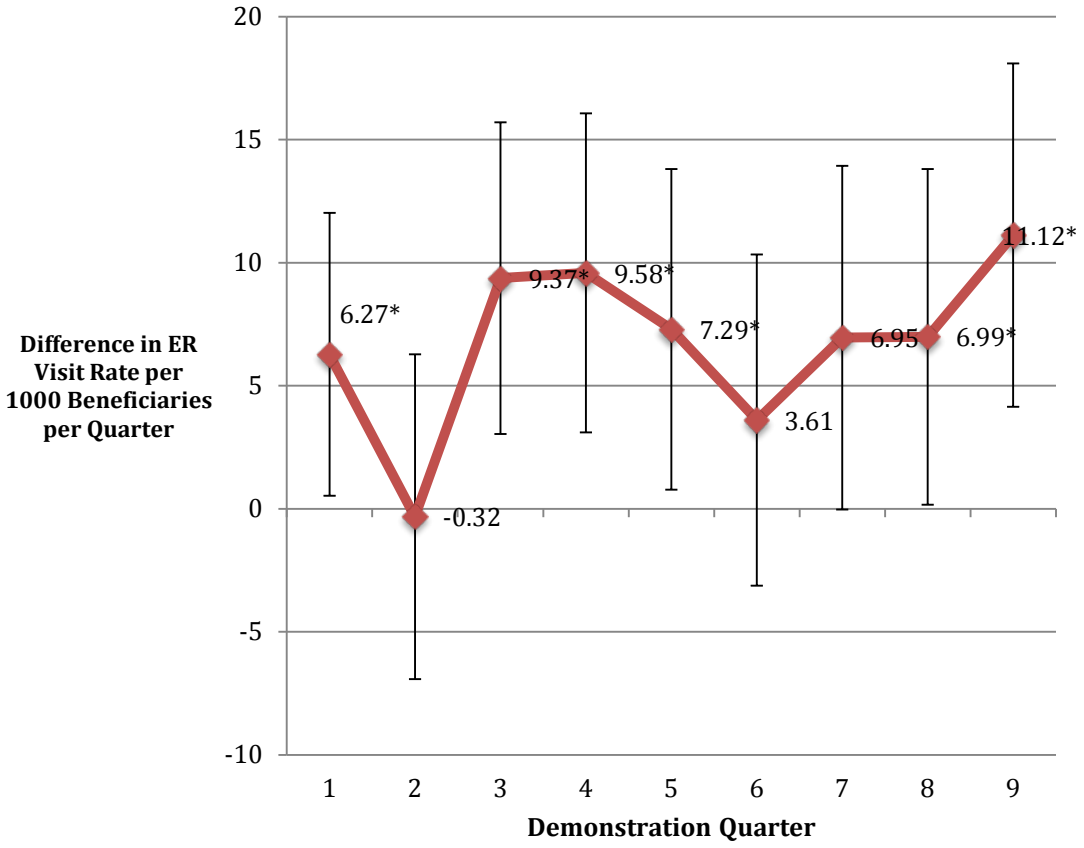
Exhibit III.14: Difference in the Proportion of Admissions with Unplanned Readmissions within 30 Days per Demonstration Quarter (Demonstration FQHCs vs. Comparison FQHCs) Based on Regression Analysis (Primary analysis: ITT Cohort Only)



ED Visits

Exhibit III.15 shows the impact of the demonstration on the number of ED visits per 1,000 beneficiaries per quarter of the demonstration period when the FQHC demonstration cohort was compared to the FQHC comparison cohort in ITT analyses. Regression results indicate that in demonstration quarters 1, 3–5, 8, and 9, there were between 6.3 and 11.1 more ED visits per 1,000 beneficiaries in the demonstration group than the comparison group (p-values <0.05). Differences between groups in all other demonstration quarters were not statistically significantly different from differences observed between groups in the baseline period.

Exhibit III.15: Difference in the Number of ED Visits per 1,000 Beneficiaries per Demonstration Quarter (Demonstration FQHCs vs. Comparison FQHCs) Based on Regression Analysis (ITT Cohort Only)



RQ2.10 How Does FQHC Participation in the Demonstration Affect Expenditures—Both Out-of-Pocket Costs (Deductibles and Coinsurance Payments) and Costs Covered Out of the Medicare Trust Fund—for Services Covered by Medicare?

RQ2.10.a Summary: Beneficiary Awareness of Costs of Seeing a Specialist (Baseline beneficiary survey)

The beneficiary survey addressed the potential effect of FQHC participation in the demonstration on expenditures with two survey items focusing on whether beneficiaries who saw specialists were aware of and/or concerned about the costs of seeing specialists. Only 20 percent of FQHC users talked with their providers about the costs of seeing a specialist.

- Demonstration FQHC users were not different from comparison FQHC users (p=0.4667).

RQ2.10.b Analysis: Beneficiary Awareness of Costs of Seeing a Specialist (Baseline beneficiary survey)

Beneficiary awareness of costs of seeing a specialist
[Analysis SQ2.3A.1a] Among beneficiary respondents who visited their attributed clinic in the last 12 months and who saw a specialist. • 18.59 percent of demonstration FQHC and 20.73 percent of comparison FQHC users reported that they *talked with their provider about the cost of seeing a specialist* – p=0.4667 for demonstration FQHC vs. comparison FQHC.
[Analysis SQ2.3A.1b] Among beneficiary respondents who visited their attributed clinic in the last 12 months and who saw a specialist, • 34.18 percent of demonstration FQHC and 39.84 percent of comparison FQHC users reported that *they were ever worried or concerned about the cost of seeing a specialist* – p=0.1025 for demonstration FQHC vs. comparison FQHC.

RQ2.10.c Summary: Expenditure Analyses (Claims data)

We assessed the impact of the demonstration on total Medicare, acute care, post–acute care, FQHC and RHC, outpatient hospital, primary care physician, and specialist physician payments. Across all demonstration quarters, comparing reported differences relative to differences observed during the baseline period:

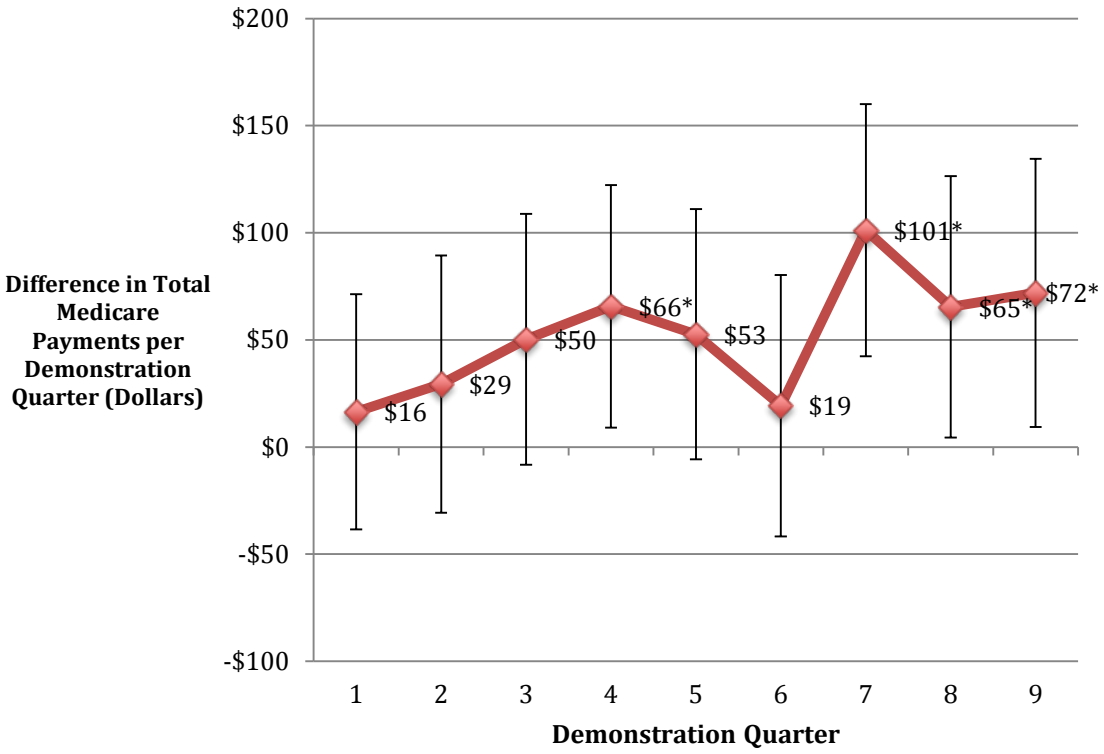
- Demonstration FQHC group users had the same or higher total Medicare payments compared with comparison FQHC users.
- Higher total payments in the FQHC demonstration group relative to the FQHC comparison group were driven primarily by differences in acute care payments in the demonstration period.

- Other payment subcategories examined were not major contributors to the overall payment increases seen in the demonstration group relative to the comparison group.

Total Medicare Payments

Exhibit III.16 shows the impact of the demonstration on total Medicare payments PBPQ for beneficiaries attributed to demonstration and comparison FQHC sites at the beginning of the demonstration (ITT cohort; analysis 1). Regression results indicate that in demonstration
quarters 4, 7, 8, and 9, total payments for Medicare beneficiaries in the demonstration group were between $65 and $101 higher than for beneficiaries in the comparison group (p-values<0.05). Differences between groups in all other demonstration quarters were not statistically significantly different from differences observed in the baseline period.

Exhibit III.16: Difference in Total Medicare Payments per Demonstration Quarter (Demonstration FQHCs vs. Comparison FQHCs) Based on Regression Analysis



RQ2.10A-B What Features Are Associated with Changes in Expenditures, and If Expenditures Change, Which Components?

RQ.2A-B.a Summary: Expenditure Components (Claims data)

Differences in total payments were driven primarily by differences in acute care payments in the demonstration period as shown in Exhibit III.17 for the ITT cohort. Acute care payments for the ITT demonstration group versus the comparison group in quarters 7 and 9 in the demonstration explain approximately half of the overall total cost increase for the demonstration group. Other subcategories for cost, summarized in Exhibits III.17–III.22 for the ITT cohort, are not major contributors to the total cost differences seen between demonstration and comparison FQHC beneficiaries in the demonstration period.

Exhibit III.17: Difference in Acute Care Payments per Demonstration Quarter (Demonstration FQHCs vs. Comparison FQHCs) Based on Regression Analysis

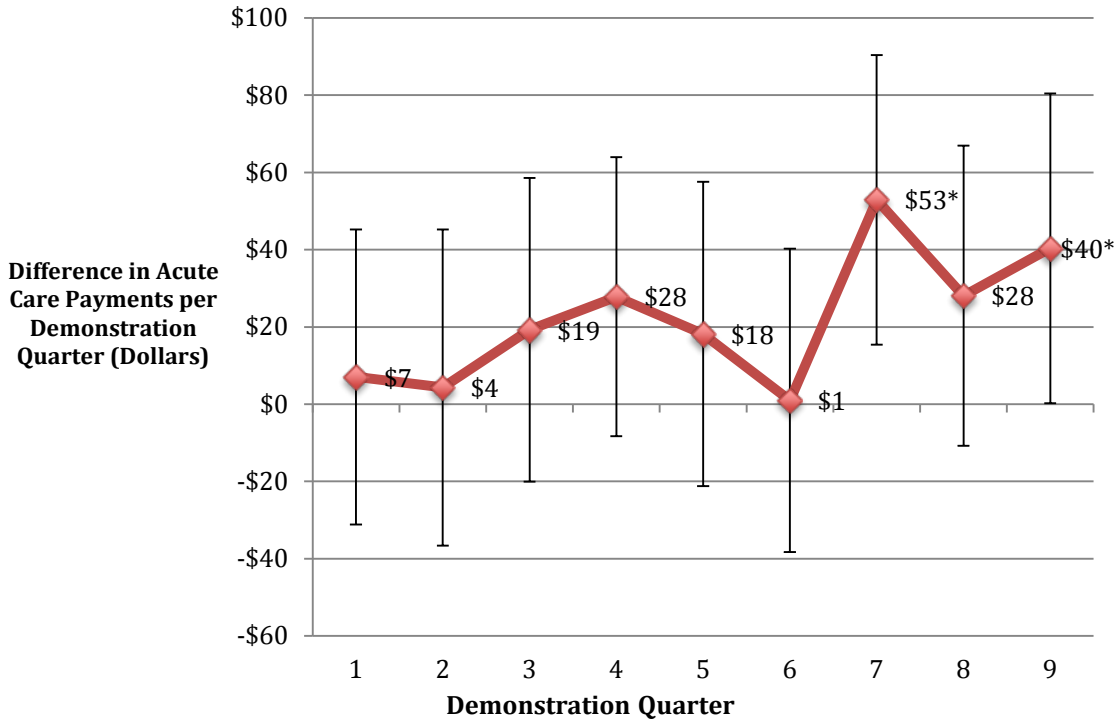


Exhibit III.18: Difference in Post–Acute Care Payments per Demonstration Quarter (Demonstration FQHCs vs. Comparison FQHCs) Based on Regression Analysis

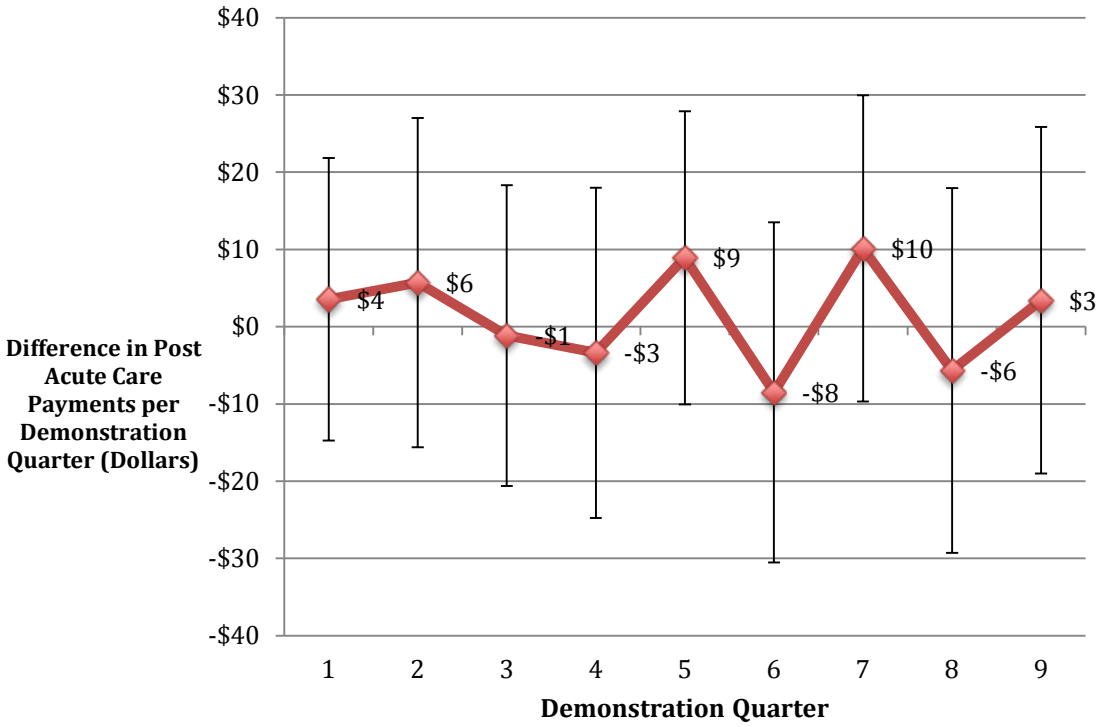


Exhibit III.19: Difference in FQHC RHC Payments per Demonstration Quarter (Demonstration FQHCs vs. Comparison FQHCs) Based on Regression Analysis

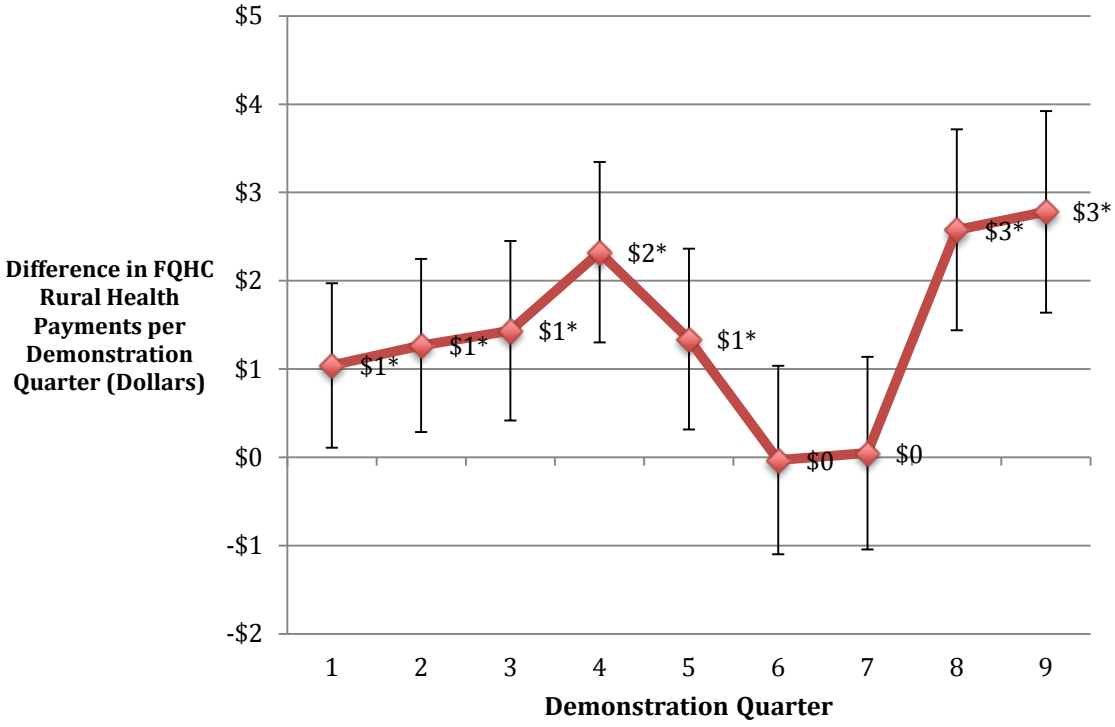


Exhibit III.20: Difference in Hospital Outpatient Payments per Demonstration Quarter (Demonstration FQHCs vs. Comparison FQHCs) Based on Regression Analysis

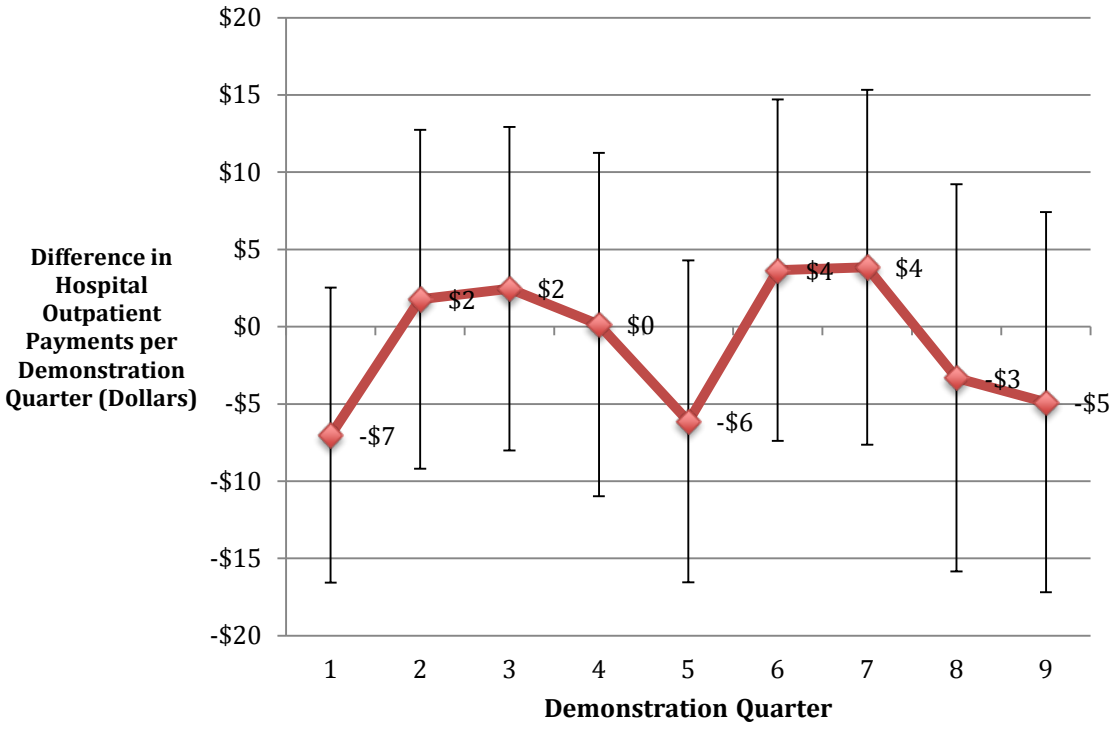


Exhibit III.21: Difference in Primary Care Payments per Demonstration Quarter (Demonstration FQHCs vs. Comparison FQHCs) Based on Regression Analysis

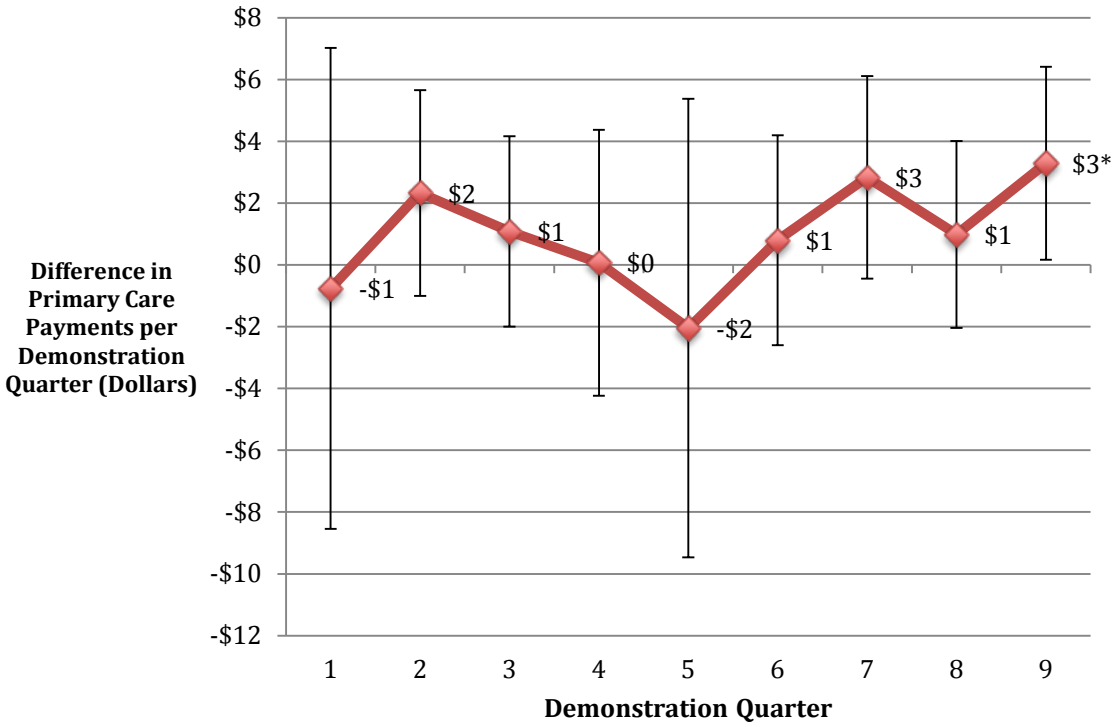
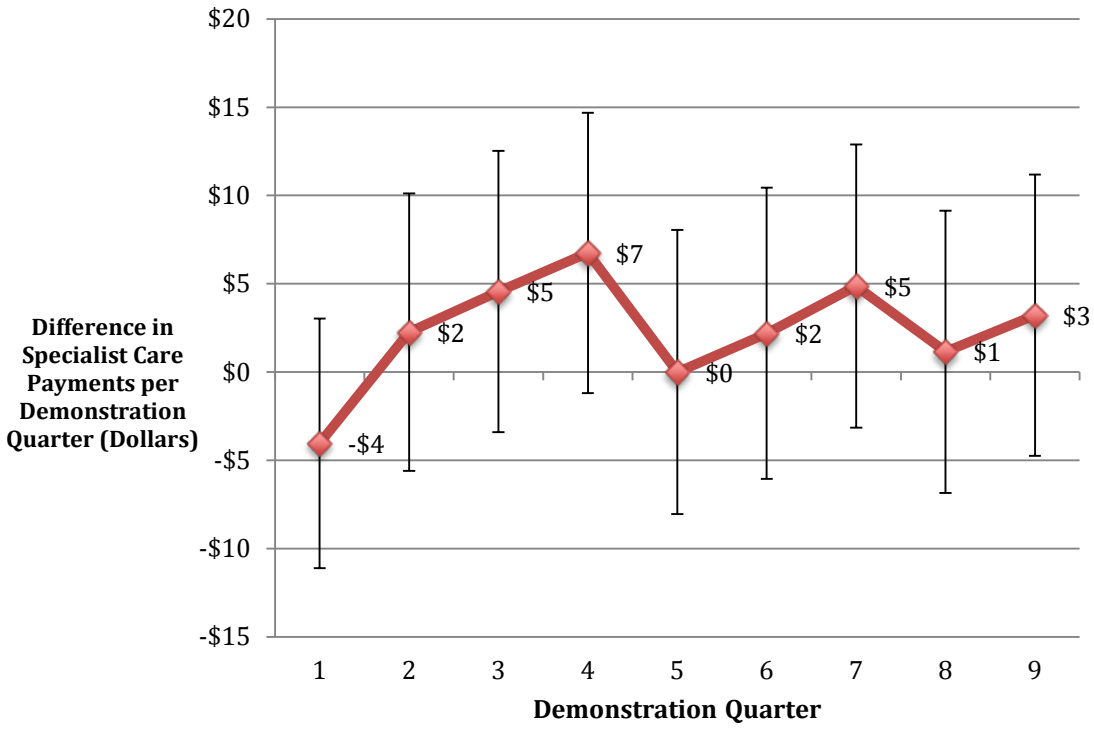


Exhibit III.22: Difference in Specialist Care Payments per Demonstration Quarter (Demonstration FQHCs vs. Comparison FQHCs) Based on Regression Analysis



RQ2.10C Is the Demonstration Budget-Neutral? If Not, Why Not? If So, How Soon into the Demonstration Are Cost Savings Seen?[27]

RQ.2.10C.a Summary: Budget-Neutral (Claims data)

Proposed Methods

The budget neutrality determination will primarily be based on analysis of the association between a beneficiary's assignment to the group of demonstration FQHCs and Medicare payments. In addition, we propose several steps that will expand upon the basic Medicare payment analyses already detailed:

1. As presented above, we propose first examining total Medicare payment differences between the demonstration and comparison groups during the

[27] The draft version of this report discussed budget neutrality in the chapter introducing Key Policy Question 3. For this final version of this report, we believe the reader will find benefit in reading about budget neutrality immediately following the discussion of expenditures. We now follow the discussion of expenditures (RQ2.10 and RQ2.10A-B) with a discussion of budget neutrality (newly numbered III.RQ2.10C; previously numbered IV.RQ3.1 and RQ3.1A).

intervention directly from the difference-in-differences model parameters. We will estimate whether these parameters indicate any statistically significant decrease in PBPQ Medicare spending—and, if so, what the magnitude of savings is. Specifically, we will examine whether any difference-in-differences model parameters are significantly less than zero, using a two-tailed t-test ($p<0.05$).

As a component of the demonstration, Medicare is paying care management fees to demonstration FQHCs at the rate of $18 per attributed beneficiary per quarter. To account for these fees in our determination of budget neutrality, we will also examine whether any PBPQ Medicare savings exceed $18 per quarter in any quarter. To do this, we will examine whether any difference-in-differences model parameters are significantly less than –$18 per quarter, using a two-tailed t-test ($p<0.05$).

If at least one of the difference-in-differences parameters is both negative and statistically significant from step (1), then we will estimate whether there is overall savings for the first such quarter through all subsequent demonstration quarters to the end of the demonstration. To determine statistical significance, we will re-parameterize our quarterly difference-in-differences model and use a two-tailed test ($p<0.05$), to determine whether aggregated payments for all quarters including and following the first cost-saving quarter are significantly less than 0 and significantly less than –$18 per quarter.

If none of the difference-in-differences parameters are both negative and statistically significant from step (1), we will further examine whether the demonstration saves costs in the following Medicare payments subcategories, using the same methods outlined in steps 1 and 3 above:

a. **Acute care payments**: This includes payments for all inpatient hospitalizations, including those to critical access hospitals.
b. **Post–acute care payments**: This includes payments for skilled nursing facilities, long-term care hospitals, and rehabilitation hospitals and distinct part units.
c. **FQHC and RHC care payments**: This includes payments for all outpatient care received at an FQHC/RHC.
d. **Hospital outpatient payments**: This includes payments for hospital outpatient visits. It excludes FQHCs, RHCs, and ED/observation beds and includes end-stage renal disease clinic visits. Laboratory and imaging payments are excluded.
e. **Primary care provider payments**: This includes payments for all primary care physicians seen outside of an FQHC or an RHC.
f. **Specialty physician payments**: This includes payments for all physician specialists seen outside of an FQHC or an RHC.

If none of the difference-in-differences parameters are both negative and statistically significant from step (1), then we will estimate the total aggregate payments over the first half of the demonstration period, over the second half of the demonstration period, and over the entire demonstration period to provide a summary of the total cost investment for CMS among demonstration sites (i.e., total payments plus care management fees) during these intervals and whether this investment was significantly greater than $0.

Preliminary Results

Steps 1–4 of budget neutrality have been conducted using the ITT analyses of Medicare payments detailed in Exhibits III.25-III.31. These analyses reveal that the demonstration is not associated with statistically significant total cost savings for CMS over any quarter of the demonstration. In fact, in quarters 4, 7, 8, and 9, the demonstration was associated with increased spending at a statistically significant level, ranging from $66 to $101 ($p<0.05$; Exhibit III.25). Expanding these preliminary overall budget neutrality analyses for the first nine demonstration quarters to the payment subcategories identified above (Exhibits III.26–III.31), these analyses indicate that the total spending increase associated with the demonstration group is driven to a large extent by increases in acute care spending (statistically significant increases in the demonstration group versus the comparison group are noted: $53 in Quarter 7 and $40 in Quarter 9; see Exhibit III.31). Other cost categories summarized are not major contributors to the total spending increase observed in the demonstration group (Exhibits III.25–III.31).

Future Work

Future work will expand these regression-based analyses using our as-treated analysis framework, where beneficiaries are attributed to demonstration or comparison sites in each quarter of the demonstration period, based on where they received the plurality of their primary care during the year prior to the start of each demonstration quarter. In this analysis, beneficiaries may enter or exit the sample and may also switch groups (demonstration/comparison) based on the outcome of the attribution process. The advantage of the described as-treated analysis is that it may more accurately reflect the costs of the demonstration because the beneficiary attribution process is updated on a quarterly basis to be more consistent with CMS's attribution process on which payments are made. (Thus, this has potential implications for cost analyses). In contrast, a potential disadvantage of this approach is that it does not allow for sample balancing methods, such as propensity score methods, since beneficiaries can be attributed to different groups in different quarters depending on where they actually received care. In addition, the difference-in-difference regression methodology proposed for our primary intent-to-treat analysis is not well suited for an as-treated analysis since the composition of the

demonstration and comparison groups is allowed to change between quarters in the baseline and demonstration periods, thus making any causal inferences from this approach unreliable. As a result, for this secondary analysis we will use a generalized estimating equation (GEE) model, where cost is modeled as a function of the changing intervention group, changing quarters, and other beneficiary, site and area characteristics, over the demonstration period only and without the added difference-in-differences feature used for our primary analysis.

RQ2.10D How Much Does It Cost to Implement and Sustain the Various Features of an APC Practice?

RQ2.10D.a Summary: APC Costs

Differing from the questions listed above that refer to Medicare payments, or costs borne by CMS, this research question refers to costs realized by the FQHC sites to implement and sustain an APCP. To fully answer this question, it would be necessary to use a micro-costing approach to collect detailed data on resources utilized to implement and sustain an APCP, as well as on the value of those resources. Such data would include the time that clinic staff spend on the additional patient-level activities that are required to be recognized as an APCP, facilities or technology acquired or maintained specifically because of their APC activities, administrative costs required to shift to and sustain an APCP, and any other materials used to deliver care through an APCP. A complete accounting of the costs of the program would combine this information on resource utilization with an appropriate cost for each item using relevant approaches, including market values and estimated costs.

Given the burden faced by FQHCs to complete achievement of NCQA Level 3 within a short period of time—and, furthermore, to successfully transform their practices—we have not identified any FQHC sites that have been willing to allow this type of data evaluation. While we will continue to gather data from site leaders about their perception of the financial incentives associated with the demonstration, the evaluation team does not anticipate being able to have adequate data to meaningfully address this question.

RQ2.10E. Is $6 Per Beneficiary Per Month (PBPM) Sufficient to Offset Those Costs? Is There a Return on Investment and, If So, How Quickly Are Costs Recouped?

RQ2.10E.a Summary: Offsets and Returns on Investment

Because of the limitations noted above in the discussion of RQ2.10D, the evaluation team is not likely to be able to address these questions quantitatively. We will continue

with our follow-up site leader interviews to query respondents about their perceptions of the adequacy of this payment.

RQ2.11 Do FQHCs Participating in the Demonstration Provide Better Coordination of Care for Medicare Beneficiaries?

RQ2.11.a Summary: Coordination of Care (Baseline beneficiary survey)

The beneficiary survey addressed coordination of care for Medicare beneficiaries using four metrics.

The first metric measures beneficiary reports of how often their specialist seemed to know the important information about their medical history.

- Demonstration FQHC users were significantly more likely than comparison FQHC users to report specialists seemed to know their important medical history (**p=0.0032**).

At this time, we do not know if the demonstration FQHC advantage is because the demonstration FQHC providers perform better outreach to specialists, are more likely to share EHRs with specialists, are more likely to have specialists on site facilitating ready access to information or select specialists more likely to outreach to FQHCs to learn the patient's history, or if this represents a different mechanism.

The second metric assesses coordination with a visit within 14 days following hospitalization between the hospitalized beneficiary and a representative of the attributed clinic.

- Demonstration FQHC users (69 percent) were more likely, following hospitalization, to report visiting or calling their attributed provider's office compared with comparison FQHC users (59 percent, p=**0.0446**).

The third metric includes two items measuring coordination, asking (1) how often the attributed provider seemed informed and up-to-date about the care they received from specialists, and (2) whether a representative from their attributed office always talks with them about the prescription medicines they are taking.

- Demonstration FQHC users were not different from comparison FQHC users ($p>0.05$).

The fourth metric assesses beneficiary reports of someone in their attributed provider's office asking them about a need for additional home health services, and about the office helping with home health services if needed.

- Demonstration FQHC users were not different from comparison FQHC users for reporting the need for home health services (p=0.3134), or for reporting their need was addressed (p=0.8377).

Care Coordination

[Analysis SQ2.4A.1c] Among beneficiary respondents who visited a specialist within the last 12 months,

- 59.41 percent of demonstration FQHC and 52.95 percent of comparison FQHC users reported their specialist always seemed to know the important information about their medical history
  - **p=0.0032** for demonstration FQHC vs. comparison FQHC (top-line p-value comparing only the always response, p=0.0921).

[Analysis SQ2.4A.3bc] Among beneficiary respondents who were hospitalized within the last 12 months,

- 69 percent of demonstration FQHC and 59 percent of comparison FQHC users reported visiting or calling their attributed provider's office within 14 days of hospital discharge
  - **p=0.0446** for demonstration FQHC vs. comparison FQHC.

Results were similar when measuring visits (regardless of calls) within 14 days following hospitalization with 65 percent of demonstration FQHC users and 53 percent of comparison FQHC users reporting such visits.

  - **p=0.0190** for demonstration FQHC vs. comparison FQHC.

[Analysis SQ2.4A.1d] Among beneficiary respondents who visited a specialist within the last 12 months,

- 56.85 percent of demonstration FQHC and 57.05 percent of comparison FQHC users reported their attributed provider always seemed informed and up-to-date about the care they received from their specialist
  - p=0.9209 for demonstration FQHC vs. comparison FQHC.

[Analysis SQ2.4A.1e] Among beneficiary respondents who visited their attributed clinic in the last 12 months,

- 82.81 percent of demonstration FQHC and 82.82 percent of comparison FQHC users reported a representative from their attributed office always talks with them about the prescription medicines they are taking
  - p=0.9953 for demonstration FQHC vs. comparison FQHC.

[Analysis SQ2.4A.2a] Among beneficiary respondents who visited their attributed clinic in the last 12 months,

- 16.17 percent of demonstration FQHC and 14.23 percent of comparison FQHC users reported a need for home health services
  - p=0.3134 for demonstration FQHC vs. comparison FQHC.

[Analysis SQ2.4A.2b] Among beneficiary respondents who visited their attributed clinic in the last 12 months,

- 17.14 percent of demonstration FQHC and 16.71 percent of comparison FQHC users reported the office provided adequate home health services to satisfy their need
  - p=0.8377 for demonstration FQHC vs. comparison FQHC.

RQ2.11.b Summary: Coordination of Care (Baseline CASE survey)

Care coordination includes information flow to FQHC sites from providers who share patients with such sites and proactive continuity of care with hospitalized patients. While the effects of demonstration participation on care coordination cannot be estimated until the conclusion of the demonstration, cross-sectional relationships between RAS levels and these dimensions of care coordination may foreshadow demonstration effects.

Baseline CASE survey analyses show that approximately 44 percent of respondents reported that hospitals usually or often notified their sites of patient admissions. Twenty-two percent reported that clinicians from their sites visited patients during their admissions, 35 percent reported that EDs notified their sites of visits, and nearly 60 percent reported that their sites received discharge summaries from hospitals and EDs. None of these dimensions of care coordination was associated with RAS level.

RQ2.12 Do FQHCs Participating in the Demonstration Provide Reductions in, or Elimination of, Health Care Disparities Among Medicare Beneficiaries?

RQ2.12.a Summary: Self-Reported Disparities (Baseline beneficiary survey)

We queried survey respondents about unfair treatment because of race, ethnicity, or English language proficiency. We found no significant difference in responses between demonstration and comparison FQHCs. As we analyze our follow-up beneficiary data, we will examine trends in measures by beneficiary subgroups. We will also pursue claims-based analyses to evaluate disparities.

RQ2.12.b Analysis: Self-Reported Disparities (Baseline beneficiary survey)

Beneficiary report of unfair treatment because of race, ethnicity, or English language skills
[Analysis SQ2.6A.1a] Among beneficiary respondents who visited their attributed clinic within the last 12 months, • 4.76 percent of demonstration FQHC and 5.37 percent of comparison FQHC users reported having been treated unfairly at this provider's office because of their race or ethnicity sometimes, usually, or always – p=0.3366 for demonstration FQHC vs. comparison FQHC.
Analysis SQ2.6A.1c] Among beneficiary respondents who visited their attributed clinic within the last 12 months and reported they did not speak English very well, • 10.74 percent of demonstration FQHC and 10.86 percent of comparison FQHC users reported *having been treated unfairly at this provider's office because they did not speak English very well* – p=0.9648 for demonstration FQHC vs. comparison FQHC.

RQ2.12.c Summary: Reduction/Elimination of Disparities (Baseline CASE survey)

Health disparities may be affected by the availability of interpreter services. While the effects of demonstration participation on disparities cannot be estimated until the conclusion of the demonstration, cross-sectional relationships between RAS levels and interpreter services may foreshadow demonstration effects.

Baseline CASE survey analyses show that more than 40 percent of respondents reported that their sites were making efforts to increase the availability of interpreter services. Compared with sites at RAS Level 1, respondents in RAS Level 2 and 3 sites were more likely to report such efforts.

RQ2.12.d Summary: Reduction/ Elimination of Disparities (Claims data)

We have not yet reached the phase of our work plan where we have analyzed a reduction of disparities using the claims-based difference-in-differences analyses. However, we will examine this during the coming year.

Analyses of Medicaid Beneficiaries

This section introduces our analyses of Medicaid beneficiaries addressing RQ 2.13.

RQ2.13 Does the Demonstration Cause Any Spillover Effects on Medicaid Beneficiaries or Other Populations Served by the FQHCs?

RQ2.13.a Summary: Spillover Effects on Medicaid Beneficiaries (Medicaid claims)

In unadjusted analyses, we observed little impact of the demonstration on utilization measures and process measures of quality, with the exception of nephropathy testing for patients with diabetes, breast cancer screening, and cervical cancer screening. In the coming months, we will implement a new method of measuring the cost of care, add data from at least two new states to these analyses, and begin to conduct adjusted analyses of the demonstration's impact.

RAND's evaluation includes an assessment of potential "spillover" effects of the demonstration on the cost and quality of care provided to Medicaid enrollees who seek care at FQHCs. Because of the added complexity of processing claims and eligibility files for individual states, we limited our focus to a small number of states. RAND initially prioritized the selection of states using the following criteria:

- number of demonstration sites and comparison FQHC sites located in the state
- quality and completeness of Medicaid encounter data (as measured by data analysts at Mathematica Policy Research)
- regional diversity
- representation of states with a high percentage of Spanish-speaking residents.

The five states meeting these criteria were California, Florida, Michigan, New Mexico, and New York.

Further exploration of Medicaid Statistical Information System (MSIS) files revealed incomplete use of NPIs during the demonstration's baseline period (November 1, 2010, through October 31, 2011). Some providers continued to use Medicaid legacy identifiers rather than NPIs even after their use became mandatory in 2007. Although we successfully cross-walked legacy identifiers to NPIs (using the most recent version of the National Plan and Provider Enumeration System database, which contains both fields) we were unable to identify any claims during the baseline period for a small number of demonstration and comparison sites. Across the top five priority Medicaid states, the percentage of sites we successfully identified ranged from 67 percent to 100 percent for both demonstration sites and comparison sites (Exhibit III.23). These findings led us to focus on California, New Mexico, and Michigan as our three highest-priority states for the analysis of Medicaid spillover effects. In the future, we may consider expanding our

analyses to additional states that were considered high priority using the criteria listed above and for which we can identify a high percentage of sites.

Exhibit III.23: Site Identification Percentage in Medicaid Claims for High Priority States

State	Number of Sites Included in RAND's Evaluation of Medicare Beneficiary Cohorts		Number and percentage of Sites Identified in Medicaid Claims			
	Demonstration Sites	Comparison Sites	Demonstration Sites		Comparison Sites	
CA	70	97	63	90%	95	98%
FL	31	47	23	74%	36	77%
MI	16	9	16	100%	9	100%
NM	25	10	25	100%	10	100%
NY	15	15	10	67%	10	67%

Our Medicaid analyses used an ITT design that followed a single cohort for a two-year period of follow-up comprising a 12-month baseline period and a 12-month demonstration period.[28] To be eligible for attribution, enrollees were required to meet four criteria for the 12-month period before the start of the demonstration:

1. age 18 or older
2. eligible for full Medicaid benefits
3. no more than a 45-day gap in eligibility per year
4. not enrolled in Medicare (i.e., non-dual enrollee).

A total of 100,283 enrollees were ultimately attributed to the 98 FQHCs we could identify in claims during the demonstration's baseline period.

Our methodology for estimating demonstration effects follows the approach we used for our Medicare cohorts. We use GEEs to model each dependent variable and estimate the demonstration's impact on outcomes for each quarter of the demonstration period. The comparison group for these analyses comprises only FQHCs because Tax Identification Numbers are not available in MSIS files. As of the writing of this report, all results reflect unadjusted analyses. We plan to begin fitting models that adjust for beneficiary, site, grantee, and area-level characteristics when we have cleaned and compiled data from additional states. In addition, we have excluded results from our analyses examining impacts of the demonstration on Medicaid costs until we implement a more robust method for imputing costs (described below).

[28] This is the same duration of time that we reported in our first annual report. Additional quarters of MSIS data submitted by California to CMS have not been approved by CMS and therefore are not available to RAND via the CMS mainframe—the mechanism by which RAND has obtained MSIS files since the beginning of the demonstration.

Demonstration Impact on Utilization Measures

In unadjusted analyses of the demonstration's impact on utilization measures, Medicaid enrollees who were attributed to demonstration FQHCs had higher rates of all-cause admissions (4.2 more admissions per 1,000 beneficiaries per quarter) in the fourth quarter of the demonstration (Exhibit III.24). Similarly, enrollees attributed to demonstration sites had an unplanned hospital readmission rate that was 3.8 percentage points higher than comparison sites in Quarter 4 of the demonstration period. In no other quarter of the demonstration's first year were there any statistically significant differences between demonstration sites and comparison sites for any other utilization measures (i.e., ED visits and admissions for ACSCs).

Exhibit III.24: Demonstration's Impact on Utilization Measures, California Medicaid Enrollee Cohort

Demonstration Impact in Demonstration Quarter	Inpatient Admission Rate per 1,000 Enrollees		ACSC Inpatient Admission Rate per 1,000 Enrollees		Hospital Readmissions		ED Visit Rate per 1,000 Enrollees	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI (%)	Estimate	95% CI
1	1.8	–1.8, 5.5	–0.3	–1.8, 1.3	–0.3	–3.5, 2.8	–8.8	–21.1, 3.5
2	0.1	–3.7, 3.9	0.0	–1.5, 1.4	0.0	–3.0, 3.1	–7.5	–20.1, 5.1
3	1.5	–2.3, 5.3	0.3	–1.2, 1.8	0.4	–2.7, 3.6	–3.9	–17.4, 9.5
4	4.2	0.2, 8.2	0.7	–0.7, 2.1	3.8	0.9, 6.7	–7.1	–21.3, 7.2

NOTE: Unadjusted estimates are reported.

Demonstration Impact on Process Measures

We found statistically significant demonstration impacts in unadjusted analyses for two of the four process measures of quality for patients with diabetes (Exhibit III.25) and for each of the two cancer screening measures (Exhibit III.26). Patients with diabetes who were attributed to demonstration sites were 1.9 percentage points less likely to receive HbA1c tests in the second quarter of the demonstration, but were 1 to 2 percentage points more likely to receive nephropathy tests in three of the four demonstration quarters. Demonstration sites provided breast cancer and cervical cancer screening exams at higher rates than comparison sites in each of the first four quarters of the demonstration—approximately 1 to 2 percentage points higher than comparison sites for each measure.

Exhibit III.25: Demonstration's Impact on Diabetes and Ischemic Vascular Disease Process Measures, California Medicaid Enrollee cohort

Demonstration Impact in Demonstration Quarter	Diabetes Process Measures						IVD Process Measures	
	HbA1c Test		LDL Test		Nephropathy Test		Lipid Test	
	Estimate (%)	95% CI (%)	Estimate (%)	95% CI (%)	Estimate (%)	95% CI (%)	Estimate (%)	95% CI (%)
1	–0.5	–1.8, 0.8	–0.6	–2.0, 0.8	1.0	0.1, 2.0	–1.7	–4.0, 0.5
2	–1.9	–3.7, –0.1	–1.9	–3.9, 0.2	1.4	0.1, 2.7	0.1	–2.9, 3.1
3	–0.2	–2.2, 1.7	–0.3	–2.6, 2.0	1.4	–0.1, 3.0	–0.1	–3.6, 3.3
4	0.8	–1.2, 2.8	–0.9	–3.4, 1.6	2.0	0.3, 3.6	0.6	–3.1, 4.3

NOTE: Unadjusted estimates are reported.

Exhibit III.26: Regression Output for Screening Measures, California Medicaid Enrollee Cohort

Demonstration Impact in Demonstration Quarter	Breast Cancer Screening		Cervical Cancer Screening	
	Estimate (%)	95% CI (%)	Estimate (%)	95% CI (%)
1	1.2	0.6, 1.8	0.6	0.2, 1.0
2	1.5	0.7, 2.3	1.2	0.6, 1.7
3	1.9	1.0, 2.9	1.9	1.2, 2.5
4	2.6	1.5, 3.7	2.5	1.8, 3.2

NOTE: Unadjusted estimates are reported.

Next Steps for Medicaid Analyses

Over the next few months, we will complete our ongoing migration from MSIS files to alphaMAX files to improve the efficiency of our data acquisition and processing. Accessing alphaMAX files via the CMS Virtual Research Data Center has greatly accelerated the speed with which we can conduct our Medicaid claims analyses. Furthermore, alphaMAX files include a greater level of processing than the raw MSIS files, including the incorporation of claims adjustments and the cross-walking of beneficiary identifiers, which also streamlines our data-cleaning procedures. Because alphaMAX files are only available from 2011 to 2013, and because our evaluation requires 2010 data for our baseline year, we have combined MAX 2010 files and alphaMAX 2011–2013 files to cover our entire evaluation period. Mathematica Policy Research is planning to release information on the comparability of MAX and alphaMAX files that will help us better assess whether any systematic bias might be introduced by combining these two datasets. RAND will conduct analyses as needed to determine whether the absence of data-cleaning "business rules" that are used to generate MAX files (but are absent from alphaMAX files) will compromise our analysis in any way. Depending on Mathematica's guidance, RAND may consider conducting a sensitivity

analysis in which we use only alphaMAX data. Such an analysis would use only the last ten months of the evaluation's one-year baseline period (January 1, 2011, through October 31, 2011) rather than a full year of baseline data. In addition, this analysis would not be able to include an assessment of the demonstration's impact on process measures because these measures typically require at least one full year of data to establish eligibility or measure the successful provision of a service.

Second, we will add more states to our analysis and begin to conduct adjusted analyses. These steps will include generating propensity score weights, assessing improvements in balance between characteristics of demonstration and comparison sites (and their attributed beneficiaries) after the introduction of weights, and assessing model fit for all regression models (including propensity score regression models). We have deferred generating propensity score weights and using adjusted analyses until we have our full cohort because covariate effects may differ in magnitude or even direction in small cohorts as compared with larger cohorts.

Fourth, we will continue to monitor the potential for incomplete managed care encounter data to introduce bias into our analyses. At a minimum, we plan to include an adjustment for enrollment in managed care in all regression models (including propensity score models) in case there are any systematic differences in quality or cost of care associated with managed care compared with fee-for-service, which might simply be a function of differences in the completeness of the underlying claims data.

IV. Key Policy Question 3: Which Practice-Site Characteristics Are Associated with Observed Changes in Structures, Including NCQA Recognition Status, and with Patient-Level Processes and Outcomes?

Exhibit IV.1: RQs and their Derivative Subquestions for Key Policy Question 3[a]

Second Annual Report Section	Research Question
IV.RQ3.1	**Are changes in beneficiary outcomes due to implementation of APCPs or to other causes?**
RQ3.1A	Is the impact of the demonstration on beneficiary outcomes mediated by the adoption of APCPs?
RQ3.1B	To what extent do demonstration sites improve beneficiary outcomes in ways that are not explained by achievement of Level 3 PCMH recognition?
RQ3.1C	What is the impact of achieving Level 3 PCMH recognition on beneficiary outcomes independent of a site's participation in the demonstration?
IV.RQ3.2	**To what extent is the demonstration's impact on beneficiary outcomes stronger for certain types of beneficiaries or sites as compared with other types?**
IV.RQ3.3	**What are the facilitators and outcomes of change?**
IV.RQ.3.4	**What are some differences and commonalities among high- and low- performing FQHCs?**
IV.RQ.3.5	**If the demonstration were to be extended or expanded, what advice might be provided to CMS?**

[a] Appendix R provides a cross walk between RQs discussed in this section and those that have previously been linked to Key Policy Question 3.

Summary of Approach to and Findings for Key Policy Question 3

Key Policy Question 1 informs us about predictors of TA uptake and how TA and other features contribute to a site's transformation to NCQA Level 3. Key Policy Question 2 asks whether demonstration sites deliver better beneficiary processes and outcomes than comparison sites. Key Policy Question 3 asks which practice and beneficiary characteristics are associated with observed changes in structures, processes, and outcomes. To transition from Key Policy Questions 1 and 2 to our final key policy question, we need additional methods to allow us to understand how differences between demonstration sites affect beneficiary outcomes. This section introduces methods that the evaluation team will study in the coming year. This chapter introduces methods that will be used to address this policy question and also presents some qualitative analyses that have already been implemented.

This section introduces mediation (Section IV.RQ3.1) as a foundation for transition from previously described analyses to this final question. With mediation, we address the question of whether changes in beneficiary outcomes are due to implementation of APCPs or to other causes. We next introduce effect modification (also in Section IV.RQ3.2) by beneficiary and practice-site attributes as an additional set of tools for approaching this final question. Effect modification tools allow us to understand the extent to which the demonstration's impact on beneficiary outcomes is stronger for certain types of sites as compared with others.

This leads to the follow-up to specific questions addressed in Section III where Key Policy Question 2 identified domains or intermediate outcomes where demonstration sites provide care to beneficiaries that differs from care delivered by comparison sites. In following up on Key Policy Question 2, Section IV.RQ3.3 asks about the facilitators of better beneficiary-level performance across the many variables analyzed in Section III and about how changes in these variables affect observed patterns in cost and utilization for demonstration sites relative to comparison sites. Though we introduced these questions and define the performance metrics for these variables in the prior chapter, Section IV.RQ.3.3 presents our approach to the facilitators and outcomes of differences between demonstration and comparison site care for beneficiaries.

Section IV.RQ3.4 asks about differences and commonalities among high- and low-performing FQHCs. We will address this with multiple approaches, including an examination of predictors of site-level advancement to higher levels of recognition (as introduced in Key Policy Question 1) and with qualitative analyses including a qualitative comparative analysis (QCA).

Finally, Section IV.RQ3.4 also explores what advice might benefit CMS should this demonstration be extended or expanded. We answer this question with thematic analyses of site leader and PCA informant interviews.

IV.RQ3.1 Are Changes in Beneficiary Outcomes Due to Implementation of APCPs or to Other Causes?

Mediation Analysis

The finding of any statistically significant and substantive demonstration impact raises the question of whether the changes in beneficiary outcomes are due to the implementation of APCPs or other causes. Answering this question is critically important not only for appropriately interpreting the results of the FQHC APCP demonstration, but also for formulating sound policies and programs based on the evaluation's findings. We developed three methodological policy questions that allow us to extend our understanding of the demonstration and the mechanisms by which beneficiary outcomes are affected through a site's participation in the demonstration and a site's achievement of Level 3 recognition. These three questions help us to refine our understanding of the mechanisms that explain differences between demonstration and comparison sites described with multiple RQs associated with Key Policy Question 2. The three questions are as follows:

- **Research Question 3.1A**: Is the impact of the demonstration on beneficiary outcomes mediated by the adoption of APCPs?
- **Research Question 3.1B:** To what extent do demonstration sites improve beneficiary outcomes in ways that are not explained by achievement of adoption of APCPs?
- **Research Question 3.1C:** What is the impact of adopting APCPs on beneficiary outcomes independent of a site's participation in the demonstration?

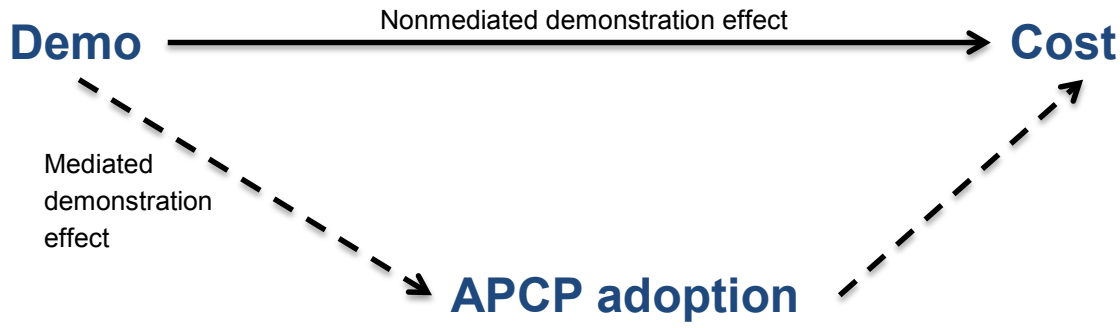
We answer these three questions using a unified modeling framework that estimates all relevant parameters simultaneously. We discuss our analytic approach in a subsequent section.

Rationale for Examining the Role of PCMH Recognition as a Mediator of Changes in Beneficiary Outcomes

While the demonstration was designed to directly stimulate the adoption of APCPs by providing TA to support practice transformation, it is possible that demonstration FQHCs may be able to improve beneficiary outcomes in ways that are not mediated by the adoption of APCPs. For example, as demonstration sites transform and greater efforts are

made to implement new care processes, some attention might be diverted from caring for the site's sickest patients in the short term as sites are fine-tuning these processes. In this example, the risk profile of the demonstration beneficiary cohort could shift to a different case mix that could be associated with lower costs. Alternately, sites might have used care management revenues to hire staff to improve their billing, rather than the anticipated use of these revenues to support APCPs. Research Question 3.1A seeks to determine whether the demonstration effect on beneficiary outcomes is mediated by the adoption of advanced primary care practices or by other factors. The *mediated demonstration effect* is indicated by dashed lines in Exhibit IV.2.

Exhibit IV.2: Demonstration Effects Estimated in RAND's Mediation Analysis



Research Question 3.1B seeks to *quantify* the extent to which the demonstration affects beneficiary outcomes through mechanisms other than the adoption of APCPs. Hereafter, we refer to this effect as the *nonmediated demonstration effect*. This question tests the demonstration's main assumptions—that most of the intervention's impact should be mediated through changes in the adoption of APCPs. Specifically, this question assesses the magnitude of the nonmediated demonstration effect in a regression analysis that adjusts for the mediator's impact on the dependent variable (i.e., the impact of Level 3 PCMH recognition on cost) and estimates the adjusted effect of the demonstration. If this effect is statistically significant and positive (i.e., outcomes are favorable), this implies that demonstration sites are able to improve beneficiary outcomes through some other mechanism than the adoption of APCPs. If the demonstration effect is statistically significant and negative (i.e., beneficiary outcomes worsen over time), it implies that participation in the demonstration may have unintended consequences—for example, quarterly submission of RAS data and attendance at webinars might distract demonstration clinicians from providing optimal care to their patients.

Research Question 3.1C seeks to measure a "medical home" effect on beneficiary outcomes.[29] This question is important because CMS and other policymakers are likely to be interested in understanding whether the PCMH model has a demonstrable effect on the cost and quality of care for Medicare beneficiaries regardless of the effectiveness of the transformation supports provided as part of the APCP demonstration. RAND has documented the slow rollout of the TA strategies during the demonstration, and in fact, the most effective TA strategies for FQHCs still may not be known. For these reasons, the intervention might not yet have affected the medical home trajectory of demonstration sites. At the same time, many FQHC comparison sites were well on their way to becoming recognized medical homes—spurred on by a concurrent recognition program sponsored by HRSA and the growth of incentive programs operated by states and commercial payers (particularly Medicaid Health Home programs). The potentially weak set of demonstration interventions combined with the potentially "exposed" comparison group suggests that our evaluation's primary demonstration impact estimates might not reflect the full effect that becoming a PCMH has on beneficiary outcomes.

The analyses supporting these questions account for the measured changes in structure that accompany the transition to a PCMH and will help to differentially estimate the impact on beneficiary outcomes from participation in the demonstration that is not explained by Level 3 recognition (nonmediated demonstration effect) and that is explained by PCMH recognition independent of participation in the demonstration (medical home effect).

Analytic Approach

These impacts will be modeled using a difference-in-differences framework similar to the overall evaluation framework but with additional information specifying the PCMH recognition status of each site. Only sites that have achieved NCQA Level 3 recognition will be designated as PCMH-recognized sites. Outcomes will be aggregated annually, and only the baseline year and the most recent year of data will be included in the analysis. For the final analysis, this will be the baseline year and the third year of the demonstration. This approach will allow for a comparison of how demonstration and comparison groups differ in the last year of the demonstration compared with the baseline year, and how PCMH-recognition sites compare to nonrecognized sites in the last year compared with the baseline year.

The specification of the difference-in-differences model will be as follows:

$$Y_{it} = \beta_0 + \beta_1 After_{it} + \beta_2 Demo_{it} + \beta_3 NCQA3_{it} + \beta_4 After_{it} * Demo_{it} + \beta_5 After_{it} * NCQA3_{it} + \beta_6 X_{it} + \varepsilon \quad (1)$$

[29] This relationship cannot be displayed using Exhibit IV.2, but is described using formula 1 below.

where:
After indicates that the outcome is collected in the last year of the demonstration (relative to baseline);
β_1 estimates the difference between the last year and the baseline period among comparison sites that have not achieved PCMH recognition by the last year of the demonstration;
Demo indicates that a demonstration site (including both current participants and sites that have attrited);
β_2 estimates the difference between demonstration and comparison sites in the baseline period among those sites that have not achieved PCMH recognition;
NCQA3 indicates that a site has achieved PCMH recognition by achieving NCQA Level 3 status by the last year of the demonstration;
β_3 estimates the difference between comparison sites that will eventually gain PCMH recognition and comparison sites that will not gain PCMH recognition, in the baseline period;
$After_{it}*Demo_{it}$ is an interaction term that allows the impact of the demonstration to vary between baseline and the last year of the demonstration for both PCMH recognized sites and non-PCMH recognized sites;
β_4 estimates the additional impact of the demonstration in the final year of the evaluation, compared with the baseline year;
$After_{it}*NCQA3_{it}$ is an interaction term that allows the difference between PCMH recognized and nonrecognized sites to differ in the last year of the demonstration, compared with the baseline year for both demonstration and comparison sites; and
β_5, estimates the impact of gaining Level 3 recognition in the final year of the demonstration.

We will also test for the statistical significance of the three-way interaction among *Demo*, *After*, and *NCQA3* that would indicate that the impact of the demonstration differs for sites with Level 3 recognition compared to nonrecognized sites in the last year of the demo when compared with the baseline year. If we find a significant interaction term that is also large in magnitude (in terms of policy relevance), we will not be able to estimate a single medical home effect or a single nonmediated demonstration effect because, by definition, each effect will depend on a site's participation status and its recognition level.

Using the parameter estimates from this model we will answer each of the three RQs for Key Policy Question 3 as follows:

Research Question 3.1A: Is the impact of the demonstration on beneficiary outcomes mediated by the adoption of APCPs? We will only seek to answer this question empirically if the demonstration effect from one or more dependent variables from our main impact analyses is statistically significant. To answer this question, we will begin with a simplified version of Equation 1 that contains only the *Demo* indicator, *After* indicator, and other adjustment variables (*X*). We will then test whether statistical

adjustment for a site's final NCQA recognition status (through the addition of the *NCQA3* indicator to the model) reduces the magnitude of the demonstration parameter estimate, and whether it causes the demonstration effect to lose statistical significance. We will conclude that NCQA recognition is a mediator of the intervention's effect on outcomes if both conditions hold.

Research Question 3.1B: To what extent do demonstration sites improve beneficiary outcomes in ways that are not explained by achievement of Level 3 PCMH recognition? We will measure the nonmediated demonstration effect using our estimate of β_4. This estimate effectively adjusts for the contribution of a site's Level 3 recognition status to the observed changes in beneficiary outcomes. In the event that our final model uses a three-way interaction term between *Demo*, *After,* and *NCQA3,* we will assess demonstration impacts within strata defined by NCQA Level 3 recognition status. Because of our large sample size, one way to operationalize this method is to fit Equation 1 separately for Level 3 sites and non–Level 3 sites, and estimate β_4 from each model. To produce estimates on the natural scale of each dependent variable, we will implement a recycled predictions methodology using the Stata margins command—analogous to the method we used to estimate our primary impact estimates. This approach will provide a recycled estimate of the nonmediated demonstration effect in Level 3 as well as in non–Level 3 sites. For a model with a three-way interaction, only the recycled differential effect of the intervention between the Level 3 and non–Level 3 can be comprehensively estimated.

Research Question 3.1C: What is the impact of achieving Level 3 PCMH recognition on beneficiary outcomes independent of a site's participation in the demonstration? We will measure the "medical home" effect using our estimate of β_5. This estimate effectively adjusts for the contribution of a site's demonstration participation status to the observed changes in beneficiary outcomes. In the event that our final model uses a three-way interaction term between *Demo, After,* and *NCQA3,* we will measure the medical home effect within strata defined by demonstration participation status. One way to operationalize this method is to fit Equation 1 separately for demonstration sites and comparison sites, and estimate β_5 from each model. We will also generate recycled estimates to enhance the interpretability of our estimated medical home effects for each outcome measure.

Limitations

These analyses have several limitations. First, Level 3 recognition is not an exogenous baseline characteristic but, in fact, one of the demonstration's main outcomes. By using a difference-in-differences approach, we assume that we can perfectly predict the baseline period in which providers will achieve Level 3 recognition by the end of the

demonstration. However, demonstration and comparison sites might differ in their propensity to become recognized in ways that are unobservable and may have cost trajectories and historical patterns in performance on quality measures that differ from sites that never become recognized—violating assumptions of the difference-in-differences model. To reduce bias, we are exploring moving to an enhanced propensity score model that seeks to match demonstration and comparison groups on time elapsed before achieving Level 3 recognition as a way to better capture the underlying trajectories in Level 3 recognition achievement while maintaining the use of our difference-in-differences framework. We recognize that the time required by sites to become recognized by NCQA may not reflect their actual rate of progression if sites chose to delay applying for formal recognition. We will continue to refine methods over the coming months to strengthen the inferences that can be drawn from all analyses of mediation effects.

IV. RQ3.2 To What Extent Is the Demonstration's Impact on Beneficiary Outcomes Stronger for Certain Types of Sites as Compared with Other Types?

The third key policy question examines to what extent the demonstration's impact on beneficiary outcomes is stronger for certain types of sites as compared with other types. We will examine this phenomenon, also known as *effect modification*, across a range of characteristics hypothesized to influence the magnitude of the demonstration's impact. These include:

- beneficiary characteristics: dual eligibility, disability, comorbidity (measured by RAND's comorbidity index)
- site characteristics: number of beneficiaries, total revenue per site, number of affiliated service delivery sites within the grantee's organization
- area-level characteristics: percentage of poverty in the census tract in which the FQHC operates, location in a rural area.

We will also examine effect modification by a site's baseline readiness for becoming a medical home, which will be measured by baseline RAS scores. Analyses using baseline RAS scores, unlike all other analyses in this section, will be limited to the subset of demonstration and comparison sites for which we have baseline RAS data.

We will test for effect modification only if the demonstration is observed to have a statistically significant and substantively large impact on beneficiary outcomes. We will test for effect modification by interacting the difference-in-differences parameter estimate by each characteristic listed above.

In addition to assessing effect modification for our main impact analyses, we will also test for effect modification associated with the two mediation effects discussed under RQ3.1—the nonmediated demonstration effect and the medical home effect. As FQHCs continue to achieve Level 3 recognition at a rapid rate, there is likely to be policy interest in better understanding heterogeneity among Level 3–recognized sites. Thus, we will apply a similar methodology to develop a profile of the type of Level 3–recognized sites that are more likely to improve beneficiary outcomes.

IV.RQ3.3 What Are the Facilitators and Outcomes of Change?

The next set of analyses seeks to better understand the demonstration impacts we observed in our analyses addressing Key Policy Question 2. These analyses comprise two types. First, we identify the factors that enable demonstration sites to achieve higher levels of performance relative to comparison sites on each of several key intermediate outcomes (See Exhibit IV.3). Second, we assess the extent to which changes in these intermediate outcomes explain the observed patterns in cost and utilization outcomes for demonstration sites relative to comparison sites over the course of the demonstration. Taken together these analyses provide additional context for interpreting the demonstration's impact.

Our approach for understanding facilitators of progress on key demonstration outcomes entails using our difference-in-differences model in conjunction with an effect modification framework. The difference-in-differences model will assess changes in each dependent variable for demonstration sites relative to comparison sites over the course of the demonstration. Predictors will include beneficiary, site, and neighborhood measures derived from both claims data and the beneficiary survey. We will interact a subset of these predictors with the demonstration impact parameters (i.e., demonstration*time terms) to estimate the degree to which each predictor moderates the demonstration's impact. Candidate predictors will be examined in bivariate analyses. Additional details on our difference-in-differences framework are available in Appendix B (Claims Methodology). Our effect modification methodology is discussed in IV.RQ3.2.

To determine the extent to which changes in the dependent variables listed in Exhibit IV.3 explain patterns in cost and utilization outcomes over the course of the demonstration, we will use a mediation framework that RAND is actively developing. In brief, this framework uses a system of equations to model the relationships between assignment to the demonstration, one or more mediators (including NCQA Level 3 recognition), intermediate outcomes (such as timeliness and access), and cost and

utilization outcomes.[30] These analyses will provide estimates of the impact of each dependent variable on cost and utilization outcomes adjusting for other potential mediators and confounders. Where CASE and qualitative analyses provide contextual data, quantitative analyses will be informed be these findings.

[30] Imai K., Keele, L., and Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, *(5)*4, pp. 309–334. Our approach is based on this method.

Exhibit IV.3: Intermediate Outcomes Associated with RQs Pertinent to Facilitators and Outcomes of Change[31]

RQ Number	Dependent Variable	Report Section Describing Dependent Variable Metrics
3.3 (1A-B)	Loyalty or continuity of care	III.RQ2.1
3.3 (2A-B)	Timely delivery of health services	III.RQ2.2
3.3 (3A-B)	Better or enhanced access	III.RQ2.3
3.3 (4A-B)	Improved access to evidence-based care	III.RQ2.4
3.3 (5A-B)	Better quality of care	III.RQ2.5
3.3 (6A-B)	More effective beneficiary, family member, and/or caregiver participation in decisions concerning care	III.RQ2.6
3.3 (7A-B)	Better self-management	III.RQ2.7
3.3 (8A-B)	Participant's experiences communicating with providers	III.RQ2.8
3.3 (9A-B)	Changes in service utilization	III.RQ2.9
3.3 (10A-B)	Changes in expenditures	III.RQ2.10
3.3 (11A-B)	Better coordination of care	III.RQ2.11
3.3 (12A-B)	Reductions or eliminations in health care disparities	III.RQ2.12

IV.RQ.3.4 What Are Some Differences and Commonalities Among High- and Low- Performing FQHCs?

RQ3.4.a Summary: Differences and Commonalities

Understanding commonalities among the high and low-performing FQHCs is a central component of this evaluation. In fact, all five of our key data sources will contribute to our approach to this question. Because the data defining high and low performance is not yet fixed (as we await data from the end of the demonstration, such as final Level 3 NCQA recognition), we do not yet have final data to address this question.

While we introduced this issue in Key Policy Question 1 with a discussion of the association between TA uptake and the achievement of Level 3 recognition/ RAS final status
(Section II.RQ1.D1), we will now supplement this question with other specifications for high performance. For example, we may supplement the dependent variable of achievement of Level 3 recognition with examination of predictors of sites achieving

[31] These RQs were previously embedded within the discussion of Key Policy Question 2. Accordingly, these RQs refer back to RQ2. For example, RQ 2.1 asks whether FQHCs participating in the demonstration provide improved loyalty or continuity of care for Medicare beneficiaries. If analyses of RQ 2.1 reveal differences between demonstration and comparison sites, then RQ 3.3 will examine facilitators (A) and outcomes (B) of those differences. Within Key Policy Question 3, we assign RQ3.3 (1A-B) to the analysis question that asks about predictors and sequelae of differences in loyalty or continuity between demonstration and comparison sites. This RQ naming convention is followed for other RQs outlined in Exhibit IV.3.

Level 3 early, or achieving a large advance in their recognition status even if they do not achieve Level 3 recognition. In Key Policy Question 1, we ask this question about demonstration sites; here, we may address this question regardless of participation in the demonstration.

We will also address the question of differences and commonalities with thematic analyses of our qualitative data sources and with a QCA. This technique is used for analyzing qualitative data sets by listing and counting combinations of variables observed in data sets including both our qualitative data (site leader interviews, PCA interviews, PCA focus groups, and beneficiary and caregiver focus groups, as well as quantitative data for these sites). We will report on this in future reports.

IV. RQ3.5 If the Demonstration Were to Be Extended or Expanded, What Advice Might Be Provided to CMS?

Here, we review feedback and suggestions for future improvements in the demonstration and PCMH initiatives from three qualitative sources. At the end of the baseline site leader interviews, we asked respondents whether they had any questions, concerns, or feedback to bring to the attention of CMS and its partners who have organized the initiative. We also asked baseline PCA leader interview respondents if they had recommendations for feedback; again, they offered a few suggested changes for the demonstration and future initiatives. We supplement these reports with recommendations for improving future PCMH technical assistance based on our previous qualitative analysis of perspectives on TA intervention components from the baseline site interviews.

These suggestions include:

- providing better coordination among a more limited number of TA providers
- examining better adaptation of the NCQA recognition standards and process to FQHCs, and other changes to support for preparing the NCQA application
- ensuring sustainability of PCMH transformation and recognition, including financial support
- increasing the time for sites to lay the groundwork and pilot PCMH changes
- starting in-person training and direct assistance as early as possible in the demonstration
- focusing and formatting feedback reports to provide actionable information for sites
- making changes to support PCAs in their roles as TA providers.

RQ3.5.a Summary: CMS Advice (Qualitative data)

Provide Better Coordination Among a More Limited Number of TA Providers

Respondents expressed appreciation for TA but also frustration about the lack of a consistent, coordinated TA effort across time.

> "The patient-centered medical home comes at us on many fronts: Joint Commission, NCQA, and other places. Everybody is doing training and wanting you to come on board. It is confusing and could be more coordinated. There is the Joint Commission and Blue Cross and NCQA. All the players could have come together and decided on the requirements together so that it counts for all of the entities, instead of all the extra work it takes for getting recognition for separate entities."

A number of respondents to the baseline site interviews emphasized how confusing it can be to have so many implementation and technical support partners, each with different sets of websites and resources. This made it difficult at times for sites to identify whom to contact for different TA needs, especially in initiatives such as the APCP demonstration, where multiple TA contractors provide many similar types of services (e.g., webinars, review of procedures, policy documentation).

> "At first, it was very confusing. It was confusing to have the AIR web portal, the NCQA recognition portal, and the CMS application website, instead of having one integrated website. With so many sites, it was hard to know whom to call to ask when there were questions. There was a learning curve in the process, which could be made smoother."

Examine Better Adaptation of NCQA Recognition Standards and Other Recognition Process Support

Specific feedback on the FQHC APCP demonstration included better adaptation of the NCQA recognition standards and process to FQHCs, as well as the following changes to other elements of the NCQA recognition process:

> "Provide coaching from NCQA, not just AIR, prior to submitting the application; the examples are not enough."

> "Provide a set of best practices [for preparing the NCQA application] that have been determined so far in the process and a forum for sharing those. We are halfway through the demonstration and those best practices would help us move quickly to meet all the accountabilities in the time frame allowed."

> "Provide feedback on next steps to take and not just a numerical score."

While several of these recommendations were ultimately implemented by AIR and their partners, site leaders expressed the view that needed interventions were

implemented far too late to support the kind of transformation that many sites might have hoped to achieve had they received more timely TA support.

In the PCA leader interviews, there was also a suggestion to provide greater access to NCQA systems to facilitate PCAs tracking the submission process of their demonstration sites, as well as improvements to the timeliness and accuracy of the information. PCA leaders also recommended greater communication between NCQA and the PCAs.

> "It would be nice to be able to see a timeline documenting where in the process the NCQA application is situated. As it moves to the different stages, if we could have access, just to see where it's sitting. That's been a struggle, because we know it might be sitting there. We don't know if it's entered yet. We don't know if it is entered and it's maybe now at Stage Two and it needs to go all the way to Stage Five.
>
> Without a concurrent understanding of the status of the NCQA recognition process, PCAs have difficulty tailoring the TA most appropriate for the sites, while sites are frustrating not knowing how to prioritize their efforts."
>
> "We're finding that [information] is not even up to date. So, we don't think NCQA is doing a great job of managing what I know is many, many applications. And their website is not up to date with where they are in the status. We just had one that if you looked at their status it said they were at Stage 3, which they've actually completed, for instance . . . As we're getting into this home stretch, that would be helpful, to have a report from NCQA so we can weight that real time, the same way that AIR was reporting on RAS and stuff like that. That NCQA communication with us would help us to prop the sites up and help them with where they're struggling."

Increase the Time for Sites to Lay the Groundwork and Pilot PCMH Changes

Two sites in the baseline interviews also made a request related to a general challenge raised by many sites on the relatively short period of time that sites were given to implement the extensive changes required by the PCMH model and recognition. As one site described:

> "More time should be allowed for being able to set up the collaborations that are necessary to meet the PCMH standards. These collaborations are hard to establish based on geographic location and resources available. It takes time to transform this large system that's been practicing this way for decades."

The second site emphasized the need for better initial self-assessment and enough time to lay the groundwork and pilot change before expecting FQHCs to be able to adopt and spread the PCMH model of care.

> "There should be a requirement that each site conduct an appropriate self-assessment before they move forward to ensure that sites have time

> to lay important groundwork, build up systems, and harness commitment before moving ahead. The magnitude of the PCMH-related changes are huge on an organization. The best approach is to work through the changes at one site and then roll out the changes across the other sites as a systemwide approach."

Consider Financial Requirements for Ensuring Sustainability of PCMH Transformation and Recognition

Respondents to the baseline site interviews mentioned concern about the sustainability of PCMH transformation and recognition without at least partial continuation of the resources and additional payments by CMS after the demonstration ends. They stressed the importance of the financial support in allowing them to maintain staff required to implement and maintain the PCMH model of care and recognition. One of these respondents also queried whether CMS could examine its reimbursement policies more broadly to better support the additional staff, patient communication, and other PCMH-related care processes not currently reimbursed under current Medicare/Medicaid payment formulas.

> "Delineate what should be the focus *after* achieving NCQA recognition. Is it cost reduction or something else?"

Start In-Person Training and Direct Assistance as Early as Possible in the Demonstration

The in-person training and direct assistance were highly valued by participating FQHCs because the person-to-person connection allowed TA to be far more relevant to the specific needs of the site. Many site leaders wished PCAs had provided or facilitated in-person training earlier in the demonstration.

NCQA in-person training (sponsored by state PCAs or attended by sites outside the APCP demonstration) and mock surveys were considered the most helpful NCQA-provided technical assistance, though availability was notably limited.

Likewise, sites highly valued in-person group trainings, one-on-one practice coaching, and mock surveys provided by PCAs. Consistently across all forms of TA, site leaders noted that earlier and more consistent TA would have been much more helpful.

Focus and Format Feedback Reports to Provide Actionable Information for Sites

Site leaders emphasized that reports that were focused and formatted to provide actionable information for sites appeared most useful in helping them progress toward goals. For example, feedback reports related to progress on specific, near-term intervention goals (e.g., RAS reports on self-reported readiness on different NCQA standards for PCMH recognition) were considered more useful by sites than those related

to longer-term outcomes (e.g., quarterly reports on health care costs and utilizations by sites' Medicare and Medicaid beneficiaries).

Make Changes to Support PCAs in Their Roles as TA Providers

Respondents in the baseline PCA leader interviews provided several suggestions that they believed would enhance the effectiveness of their roles as TA providers. These included ongoing training to PCA practice coaches, consistent and predictable funding for TA through the demonstration, and streamlining the administrative requirements of PCAs.

Maintaining Training to PCA Practice Coaches

One PCA noted the value of the training for practice coaches, which had stopped at the time of the interview, and requested that AIR provide that training in lieu of other national partners no longer involved in the demonstration.

> "I think with AIR, if they can start providing more training for the coaches, that would be helpful. It kind of stopped. I would love to see real, continued training for the coaches at a coach level."

Avoiding Changes to Funding for TA Midway Through the Demonstration

PCAs noted frustration that midway through the demonstration, they learned that 20 percent of PCA funding was going to be conditional on meeting certain performance benchmarks. Although the PCAs did not make any specific requests to reverse or amend the change, they indicated the perceived need to budget their effort on 80 percent of the funding since the remaining amount could not be guaranteed.

> "We were disappointed in the way the funding came down; we didn't get as much funding in year two. We didn't know that they were going to do an 80/20, where if you don't achieve certain benchmarks, then you don't get the entire funding. So basically we do our budget on the 80 percent because you can't guarantee that you're going to get anything else. None of that was identified up front. And, so, that was disappointing."

Streamlining Support for PCAs

Some PCAs expressed concern about the amount of time they were expected to participate in conference calls.

> "I just think the phone calls are hard. It's hard to get people on board when there seems to be a phone call every other week. I could be wrong with the time, but it seems to me like there's a lot of calling and I'm not sure it's necessary to call in that much . . . It may be because we're lead PCA; we get lead calls and PCA calls and then we have calls . . . when the ancillary people are supposedly on the call, we're on that call. So it's a lot more."

> "Frequency is one thing, [duration is another]. It would be possible to maybe cut it from an hour to half an hour. Some of the calls necessitate an hour, maybe even longer. But, if there's an opportunity to shorten the call, I know that our sites would appreciate [shorter calls] especially given the demands on their time."

> "I think sometimes we're so caught up in all of these different requirements to be in our phone calls or webinars or whatever, and nobody has any time left to do the work that needs to be done. And I think that that's kind of what the health centers are feeling, too. So now we have to have a phone call with them on a monthly basis, and it's just adding more stuff to it."

Future reports will update these analyses based upon our final round of qualitative assessments, as well as observations and lessons learned from longitudinal analyses of claims, beneficiary surveys, and CASE surveys.

V. Conclusion

The APCP demonstration supported by CMS has been an enormous undertaking. The three-year demonstration aimed to support FQHCs and the beneficiaries they serve by enhancing efforts to achieve APC. Participating FQHC sites were expected to obtain Level 3 PCMH recognition within the span of the three-year demonstration, motivated and supported by four components of the intervention. Thus far, we have not found evidence that the first of these components, quarterly care management payments from CMS to participating FQHCs, has affected the transformation toward APC attributes. Qualitative analyses indicate that all interviewed site leaders appreciated the payment provided. None committed to a statement that the payment was adequate for the task; several noted that transformation is a process that continues for years. They did not know what the cost of transformation would be, but they noted concerns that the three-year duration of the intervention's fiscal payment component was too short to allow them to reach the transformation that they hoped to achieve. We will re-evaluate this question after accessing final end-of-the-demonstration NCQA recognition data.

TA from multiple sources—NCQA, AIR, PCAs, Qualis and others—represented two additional components of the intervention. In aggregate, these TA modalities were expected to present a multidimensional TA approach to a diverse set of FQHC sites that varied enormously with respect to their understanding of (and capabilities for) transformation. Despite substantial planning and implementation, the approach toward TA began slowly, in a series of fits and starts that spanned almost two years of the three-year demonstration. The multiple course corrections in TA protocols during the first half of the demonstration were responsible for some of the initial motivation of sites. This was especially true as sites were pulled in several different directions while they tried to simultaneously submit a successful application for achievement of NCQA Level 3 recognition and transform their practices with new methods of delivering care for patients.

The fourth component of the intervention was feedback reports to sites documenting their performance and improvement. While some sites acknowledged the value of these reports, many indicated they were too busy to access them.

At the same time that demonstration sites took advantage of transformation opportunities, so did comparison sites, raising a new challenge to the evaluation by exposing the FQHC comparison sites to demonstration-like interventions.

Overall, the demonstration was ambitious and many aspects of the interventions have been delivered persuasively by seasoned experts knowledgeable in practice

transformation and motivational techniques. Nevertheless, the slow start of an integrated TA program almost certainly affected sites' opportunities to take advantage of TA, likely the demonstration's main intervention component. The changes in course for the TA plans also affected the evaluation, since each new direction in TA was associated with a different protocol for tracking the sites' exposure to and uptake of TA.

Of three key policy questions addressed by RAND's evaluation team, the first pertains to the impact of the demonstration components on achievement of NCQA Level 3 recognition. The final end-of-demonstration recognition status has not yet been determined and sites have several reasons to delay their submission of the application until the very end of the demonstration. Since they receive no "credit" for recognition less than Level 3, they have incentive to delay their application until they are confident they will succeed (or at least have used all available time in an effort to succeed). Additionally, there is an expiration date on NCQA recognition, further motivating sites to wait as long as possible to apply, thereby delaying the expenditures for reapplying once the status expires.

For all of these reasons, many sites that may be capable of achieving Level 3 recognition did not apply for recognition until after the close of the analysis period for this report. This has created challenges for the evaluation team, since this central marker of demonstration success is not a reliable means for discriminating sites with more APC attributes if many sites have not applied for it. In fact, to date, the evaluation team has not yet identified a significant link between uptake of TA and achievement of Level 3. Once the final Level 3 recognition status for the demonstration sites is assigned, the evaluation team will repeat these analyses.

Our second key policy question asks whether, with time, demonstration site changes favorably affect outcomes and costs more than comparison site changes do. We ask this question focusing on a diverse set of clinically meaningful processes, as well as outcomes including utilization and cost. We have included multiple data sources to assess these variables, including claims data, beneficiary self-reports, clinician and staff reports, and qualitative interviews and focus groups with site leaders, PCAs, and beneficiaries and their families. We do have longitudinal claims data, but our beneficiary survey, CASE survey, and qualitative follow-up data collection activities are currently in the field. Accordingly, even our early longitudinal analyses are not yet complete as we are awaiting multiple forms of longitudinal data. While we await these forthcoming data, we have used every opportunity to refine our designs and to test our methods. During this coming year, we will apply our methods to more-complete data sets and merge multiple data sources allowing us to optimize multimethod analyses.

To date, our analyses have indicated higher Medicare costs and utilization in demonstration FQHCs than at comparison sites. We do not yet know whether the

observed additional costs for demonstration sites are associated with better (or worse) clinical outcomes and beneficiary experiences for FQHC users. That can only be determined at the end of the demonstration, when we have additional longitudinal beneficiary, clinician and staff, and claims data with contextual qualitative analyses that will be critical to our completing this evaluation.

This year, the evaluation team will address the third key policy question pertinent to the practice-site and beneficiary characteristics associated with changes in structures, processes, and outcomes. This report presents advances in methodologies, successful evolution of our data sets, and the presentation of an enormous body of information describing the demonstration and the characteristics of the sites, clinicians, staff, administrators, and beneficiaries associated with it.

The evaluation team looks forward with enthusiasm to this next year of work. We anticipate we will have adequate data, methods, and time to meaningfully address the key policy evaluation questions.

Preparing for the Final Evaluation Report

This second annual report presents analyses that answer some of the many RQs that support our three key policy questions. During the coming year, our new access to linked longitudinal data will allow us to more fully utilize the methods we have been using (for example, applying difference-in-differences analyses to beneficiary survey data as a supplement to claims). The linking of multiple data will allow us to move through the sequence of our policy questions and finally provide more comprehensive responses to Key Policy Questions 1, 2, and 3.

At the time of the writing of this report, the demonstration is coming to a close and the many stakeholders who have participated in and been affected by the demonstration are sharing their stories. Through site visits, longitudinal site leader interviews, PCA interviews, PCA focus groups, and beneficiary and caregiver focus groups, we have been learning how clinics aspire to and sometimes do achieve NCQA recognition. We are hearing the stories of those who are celebrating success and those who are frustrated by a lack of fiscal support, personnel, or leadership. We are learning about the experiences of beneficiaries, clinicians, and staff. We have recently completed our fielding of the final CASE survey—and during the coming months, we will access our final quarters of claims data and complete our final cohort of late (follow-up) beneficiary surveys. While we received our final RAS data in spring 2014, we anticipate receiving the final demonstration-associated NCQA recognition data for demonstration and comparison sites during winter 2015. With the influx of all of these data sets, we will, for the first time, have longitudinal data available across multiple modalities. As planned, the final year of

our analysis will involve completing our longitudinal analyses and focusing on our linkage analyses.

While this second annual report presents research findings that have been addressed by more than one type of data, the final report will synthesize these data sources, allowing us to tell the evaluation story as planned. Since we will have complete NCQA Level 3 recognition data, we anticipate that Key Policy Question 1 analyses will more clearly articulate the predictors of Level 3 recognition. Our regression methodologies will be supplemented by qualitative comparative analyses, and these will be supplemented by thematic analyses, each exploring why some sites achieve Level 3 recognition early, some late, and some not at all.

The addition of longitudinal beneficiary data will enhance claims analyses of Key Policy Question 2. For the first time with this evaluation, we will be able to tell the stories of beneficiaries over time with breadth, using claims and survey data. This integrated data set will serve as important variables for assessing patient experiences. CASE survey data and RAS/NCQA data will provide important variables to allow us to understand how practices change and the impact of practice change on beneficiaries. Qualitative analyses will enrich our understanding of these data and help us to prioritize our inquiries.

We anticipate being able to understand whether the observed higher costs for demonstration versus comparison FQHCs relative to baseline differences is associated with differences in beneficiary experiences. We also anticipate studying across multiple domains, how, if at all, beneficiary experiences change with the demonstration, and whether observed changes differ for those receiving care in demonstration sites compared with other sites.

We will be able to shed further light on costs and utilization. For example, we want to know whether improved outcomes are associated with higher total costs and hospitalization rates overall or only among the sickest or for a different beneficiary subgroup.

Further, we anticipate being able to distinguish which changes in beneficiary experience relate to the achievement of NCQA Level 3 compared with other possible explanations. All of these findings will newly be included in the final report, along with analyses to help us supplement the story of this demonstration with lessons learned that can inform future demonstrations and effect change.

Appendix A. Data Sources

Data Sources Used to Create This Second Annual Report:

- **Census data:** Census tract–level characteristics were derived using five-year aggregated data from the American Community Survey (2005–2009).
- **Claims and enrollment data:** National Claims History "TAP" files consisting of quarterly extracts of Medicare Parts A and B claims and enrollment data from the Actuarial Research Corporation for every beneficiary who has at least one visit to a federally qualified health center (FQHC). The claims data for the last reporting quarter include four months of runout and, therefore, may not include all final-action claims. Other claims have been refreshed for a 12-month runout. Claims-based utilization and cost measures are available through the demonstration's sixth quarter.
- **Claims and enrollment data for Medicaid files:** Quarterly Medicaid Statistical Information System (MSIS) files from the period November 2010 through January 2013 were obtained by RAND from the Centers for Medicare and Medicaid Services (CMS) mainframe. Claims files include original claims and any adjustments to original claims as separate records.
- **CMS payment data:** The amount paid by CMS via each payment contractor to FQHCs participating in the demonstration.
- **RAND attrition tracking:** RAND compiles information provided by CMS on FQHCs dropping out or excluded from the demonstration, as well as late entrants (current through July 17, 2014).
- **Uniform Data System (UDS):** Health Resources and Services Administration (HRSA) data containing characteristics of all Section 330 grantees, including grantee-level clinical measures, patient demographics, number of user visits by diagnosis, revenue sources, staffing information, and accreditation information. The UDS began including FQHC "look-alikes" during calendar year 2012.
- **Readiness Assessment Survey (RAS):** A self-assessment completed by an FQHC that includes questions assessing progress toward becoming a patient-centered medical home (PCMH). These data are available through May 2014.
- **National Committee for Quality Assurance (NCQA) PCMH Recognition Status:** Truven, CMS' implementation contractor, provides the NCQA PCMH Recognition level achieved by each demonstration FQHC, including the date when recognition was achieved, through September 24, 2014.

- **Clinician and Staff Experience (CASE) surveys:** Data analyses were derived from the baseline demonstration site clinicians and staff surveys fielded May 14, 2013, through September 1, 2013.
- **Site leader interviews:** These are qualitative analyses from 20 demonstration and ten comparison FQHC site leader interviews conducted during the summer and autumn of 2013.
- **Primary care associations (PCAs):** Qualitative analyses of interviews with lead representatives of all six PCA regions conducted during August-October, 2013 and analyses of PCA focus groups (February-March 2014).
- **Technical Assistance (TA) Participation Reports:** American Institutes for Research (AIR) provides information on FQHCs participating in each TA seminar through July 31, 2014, and CMS's contractor, Truven, provides information on FQHCs participating in webinars provided by NCQA.
- **Baseline beneficiary survey:** This survey was fielded on Medicare beneficiaries attributed to a demonstration FQHC, comparison FQHC, or comparison primary care clinic (PCC) from May 23, 2013, through March 18, 2014.

Appendix B. Medicare Claims Methods

We used the plurality rule to attribute Medicare beneficiaries to a provider's office based upon the provider who offers the greatest number of primary care services over the 12-month period that immediately preceded the demonstration. Using this attribution methodology, we assigned beneficiaries to demonstration FQHC, comparison FQHC, or PCC cohorts. RAND's primary analysis uses an intention-to-treat (ITT) design in which beneficiaries were attributed to demonstration or comparison sites based on utilization during a look-back period (November 1, 2010, to October 31, 2011), defined to be one year prior to initiation of the demonstration. These beneficiaries were then followed for the remainder of the demonstration and all quality, utilization, and cost outcomes associated with these beneficiaries were assigned to the site to which the beneficiary was first attributed. RAND's analyses indicate that Medicare beneficiaries who are FQHC users exhibit strong loyalty to a single FQHC—a finding that is supported by high rates of repeated attribution to the same FQHC on a quarter-to-quarter basis, and an extremely low rate of switching between demonstration sites to comparison sites or vice versa.

Medicare Claims Cohorts

We use an ITT analytic design for four different evaluation analyses shown in Exhibit B.1. Our primary analysis includes an ITT-only cohort attributed at baseline that compares beneficiaries attributed to demonstration FQHCs with beneficiaries attributed to comparison FQHCs. Secondary analysis 1 also includes beneficiary cohorts attributed at baseline but compares demonstration FQHCs with comparison PCCs rather than comparison FQHCs. Secondary analyses 2 and 3 each involve a larger cohort of beneficiaries composed of those attributed at baseline supplemented by those who are attributed after the initiation of the demonstration. Secondary analysis 2 defines the comparison group as beneficiaries from comparison FQHCs, while secondary analysis 3 defines the comparison group as beneficiaries from comparison PCCs. We intend to examine all four of these analyses separately. For each unique cohort, we evaluate the impact of the demonstration by comparing processes, utilization, or cost outcomes for beneficiaries associated with demonstration sites with those from either FQHC or PCC comparison sites.

Exhibit B.1: Cohorts and Site Types Comparing the Four Evaluation Analyses for an Intention-to-Treat Analytic Design

		Site of First Attribution	
	Cohort(s)	Demonstration	Comparison
Primary Analysis	Attributed at baseline	FQHC	FQHC
Secondary Analysis 1	Attributed at baseline	FQHC	PCC
Secondary Analysis 2	Attributed at baseline + late entrants	FQHC	FQHC
Secondary Analysis 3	Attributed at baseline + late entrants	FQHC	PCC

Characteristics of Medicare Claims-Based Beneficiary Cohorts

This section provides an overview of the ITT cohort and the late entrant cohort that will be reported in future reports.

Intention-to-Treat Cohort. The ITT cohort includes 902,406 beneficiaries—152,300 of whom are attributed to FQHC demonstration sites, 275,847 of whom are attributed to FQHC comparison sites, and 474,259 of whom are attributed to PCC comparison sites (Exhibit B.2).

Nearly half of the ITT cohort that is attributed to demonstration and comparison FQHCs are younger than age 65, and just over half are disabled (52.3 percent and 51.2 percent for demonstration and comparison sites, respectively). Just under half of beneficiaries attributed to demonstration and comparison FQHCs are dually eligible for Medicaid (49.4 percent and 47.7 percent, respectively). The majority of beneficiaries across all groups are white; however, Asian beneficiaries are nearly twice as likely to be attributed to demonstration FQHCs than comparison FQHCs (4.1 percent vs. 2.2 percent). The cohort of PCC users differs notably by being far less likely to be disabled (41.7 percent) and dually eligible for Medicaid (37.4 percent). PCC users tend to be older, are more likely to be white, and more than twice as likely to be living in nursing homes. The most common clinical conditions among ITT cohort members are diabetes, severe mental health disorders, chronic lung disorders, cardiovascular disorders, and chronic heart failure. FQHC users were more likely than PCC users to have severe mental health disorders (15 percent vs. 11 percent), but were less likely to have heart or vascular conditions (by approximately 3 to 5 percentage points)—consistent with the older age of the PCC cohort.

Late Entrant Cohort. The late entrant cohort broadly resembles the ITT cohort with a few notable exceptions. Beneficiaries who enter the demonstration after baseline are more likely to be younger than age 65 (by approximately 20 percentage points for FQHC users and by 6 percentage points for PCC users). These beneficiaries are slightly more likely to be disabled and slightly less likely to be dually eligible for Medicaid (by about 3

percentage points each). Late entrants are more likely to have mental health conditions (by about 2–3 percentage points across FQHC and PCC users) and less likely to have diabetes (by about 2–5 percentage points).

Characteristics of Medicaid Claims-Based Beneficiary Cohort

The Medicaid claims-based beneficiary cohort is currently limited to an ITT-only cohort including beneficiaries who are attributed to demonstration and comparison FQHCs operating in California in the year preceding the demonstration. In the coming months, these analyses will be expanded to include late-entrant cohorts and will include at least two additional states, as explained in the discussion of the Medicaid analyses.

The California claims-based Medicaid cohort comprises 100,283 patients—30,579 of whom are attributed to FQHC demonstration sites and 69,704 of whom are attributed to FQHC comparison sites (Exhibit B.3). The cohort is much younger than the Medicare beneficiary cohorts—with nearly half of the cohort being younger than 40 years of age. Compared with the Medicare cohorts, the California Medicaid cohorts have a much higher percentage of women (72 percent vs. 56 percent) and a much greater representation of Hispanic and Asian patients—reflecting the percentages of these groups in the general population in California compared with other states. California demonstration sites are more likely to enroll a larger share of Asian patients compared with comparison sites (11.1 percent vs. 3.4 percent), and slightly less likely to enroll Hispanic patients (38.9 percent vs. 44.5 percent) compared with comparison sites. Both groups had similar percentages of managed care enrollees (nearly 80 percent of the cohort) and similar percentages of enrollees with disabilities (32 percent). Diabetes, chronic and disabling mental health conditions, and neurological disorders are the most prevalent clinical conditions within the Medicaid cohort.

Characteristics of Demonstration and Comparison Sites

The evaluation analyses included 503 FQHC demonstration sites, 827 FQHC comparison sites, and 1,083 PCC comparison sites. Comparison sites were selected using a propensity score–matching algorithm as described in the first annual report.

The three types of sites were equally likely to be located in rural areas (64 percent) and were well balanced in terms of the racial composition of the census tracts in which individual sites were located (Exhibit B.4). Demonstration and comparison sites were distributed across six geographic regions. Sites were most likely to be located in the Central region and, among FQHCs, least likely to be located in the Northeast. Demonstration and comparison FQHCs had comparable numbers of each of eight categories of clinicians on staff, on average, while PCC sites were more likely to have smaller numbers of primary care clinicians and midlevel providers. In addition, PCCs had

approximately twice as many clinicians on staff, on average, as FQHCs. Across the three types of sites, very few had NCQA PCMH recognition at baseline (according to the 2008 standards); however, FQHC demonstration sites were more likely than FQHC and PCC comparison sites to have achieved Level 3 recognition (5.8 percent vs. 2.7 percent and 1.0 percent, respectively).

While the demonstration started with 500 sites, 64 FQHCs are no longer participating in the demonstration and three were enrolled as late entrants (Exhibit B.5). The most common reason FQHCs ceased participation prior to July 2014 was disqualification by CMS for no longer meeting demonstration requirements (e.g., failure to complete the RAS; no longer operating as an independent site; below beneficiary count threshold). On July 3, 2014, CMS terminated 27 sites for failing to progress toward Level 3 recognition. The number of currently participating FQHCs is 439 (500 FQHCs in the demonstration–64 terminations+3 late entrants).

Among FQHCs only, most sites have been in operation for a period longer than five years, and both groups have similar distributions of duration of operation. FQHC sites vary in the size of their parent FQHC organization (as measured by the number of service delivery sites). Demonstration FQHC sites tend to be affiliated with larger FQHC organizations—43 percent of sites are affiliated with FQHCs that operate 11 or more delivery sites as compared with only 27 percent of comparison FQHC sites. Very few sites represent single-site FQHC organizations. A total of 21 sites participated in the Safety Net Medical Home Initiative—12 demonstration sites and nine comparison sites. Demonstration FQHCs were more likely than comparison FQHCs to have applied to HRSA for supplemental funding to support PCMH transformation (93.6 percent vs. 67.7 percent in fiscal year 2011), and also more likely to have received Affordable Care Act (ACA)–related grants to enable expansion and infrastructure improvements. Because both HRSA and ACA grants were disbursed to the FQHC organization rather than to individual sites, it remains unclear to what extent individual demonstration or comparison sites benefited from these awards.

Our Medicaid claims analyses included 56 demonstration FQHCs and 92 comparison FQHCs operating in California (Exhibit B.6). Demonstration sites were more likely to be located in rural areas than comparison sites (51.8 percent vs. 42.4 percent) and were more likely to be located in neighborhoods with a higher density of Asian residents, while comparison FQHCs were more likely to operate in areas with a greater number of Hispanic residents. Comparison sites included in these analyses were slightly more likely to have been operating for longer periods of time. No demonstration sites and only a single comparison site had achieved Level 3 PCMH recognition (2008 NCQA standards) at baseline. All other characteristics were well balanced between demonstration and comparison FQHCs and followed similar patterns to the larger sample of sites included in the Medicare claims analyses.

Exhibit B.2: Characteristics of Medicare Claims-Based Beneficiary Cohorts Based on Site of First Attribution

	ITT Cohort			Late Entrant Cohort*		
Characteristics	**FQHC Demonstration Sites (n=152,300)**	**FQHC Comparison Sites (n=275,847)**	**PCC Comparison Sites (n=474,259)**	**FQHC Demonstration Sites (n=102,499)**	**FQHC Comparison Sites (n=163,334)**	**PCC Comparison Sites (n=290,365)**
Age: < 65 years, n (%)	69,296 (45.5)	121,665 (44.1)	145,689 (30.7)	66,591 (65.0)	108,207 (66.2)	104,689 (36.1)
65–74 years	52,016 (34.2)	95,600 (34.7)	166,473 (35.1)	23,132 (22.6)	36,347 (22.3)	109,330 (37.7)
75–84 years	23,618 (15.5)	44,164 (16.0)	112,213 (23.7)	9,738 (9.5)	14,717 (9.0)	52,879 (18.2)
85+ years	7,370 (4.8)	14,418 (5.2)	49,884 (10.5)	3,038 (3.0)	4,063 (2.5)	23,467 (8.1)
Gender: Male, n (%)	67,714 (44.5)	121,355 (44.0)	200,298 (42.2)	47,116 (46.0)	75,462 (46.2)	120,641 (41.5)
Female	84,586 (55.5)	154,492 (56.0)	273,961 (57.8)	55,383 (54.0)	87,872 (53.8)	169,724 (58.5)
Race/Ethnicity: White, n (%)	105,286 (69.1)	192,079 (69.6)	359,571 (75.8)	69,970 (68.3)	111,100 (68.0)	222,865 (76.8)
Black	26,178 (17.2)	50,794 (18.4)	69,626 (14.7)	18,003 (17.6)	31,481 (19.3)	43,087 (14.8)
Asian	6,272 (4.1)	6,022 (2.2)	12,193 (2.6)	3,485 (3.4)	3,357 (2.1)	5,312 (1.8)
Hispanic	9,796 (6.4)	19,707 (7.1)	18,693 (3.9)	7,593 (7.4)	12,285 (7.5)	11,189 (3.9)
North American Native	1,971 (1.3)	2,328 (0.8)	6,496 (1.4)	1,104 (1.1)	1,400 (0.9)	3,065 (1.1)
Other/Unknown	2,797 (1.8)	4,917 (1.8)	7,680 (1.6)	2,344 (2.3)	3,711 (2.3)	4,847 (1.7)
Disabled, n (%)	79,655 (52.3)	141,168 (51.2)	197,992 (41.7)	56,464 (55.1)	89,423 (54.7)	129,677 (44.7)
Dual eligible, n (%)	75,292 (49.4)	131,449 (47.7)	177,494 (37.4)	46,777 (45.6)	72,343 (44.3)	100,351 (34.6)
Nursing home resident, n (%)	2,282 (1.5)	4,952 (1.8)	17,455 (3.7)	2,410 (2.4)	2,539 (1.6)	13,724 (4.7)
Clinical conditions: HIV/AIDS, n (%)	1,970 (1.3)	3,459 (1.3)	3,139 (0.7)	1,354 (1.3)	1,863 (1.1)	2,015 (0.7)
Autoimmune disorders	6,625 (4.3)	12,127 (4.4)	27,356 (5.8)	4,426 (4.3)	6,828 (4.2)	18,998 (6.5)
Severe hematological disorders	1,201 (0.8)	2,171 (0.8)	4,778 (1.0)	493 (0.5)	726 (0.4)	2,055 (0.7)
Chronic lung disorders	25,034 (16.4)	45,777 (16.6)	87,891 (18.5)	15,032 (14.7)	23,400 (14.3)	53,984 (18.6)
Cancer	12,782 (8.4)	23,643 (8.6)	49,827 (10.5)	7,571 (7.4)	11,923 (7.3)	30,734 (10.6)
Chronic alcohol/other drug dependence	6,717 (4.4)	10,603 (3.8)	14,116 (3.0)	6,067 (5.9)	9,016 (5.5)	14,459 (5.0)
Chronic mental health conditions	24,445 (16.1)	40,982 (14.9)	50,304 (10.6)	19,391 (18.9)	28,232 (17.3)	39,940 (13.8)

Characteristics	ITT Cohort			Late Entrant Cohort*		
	FQHC Demonstration Sites (n=152,300)	FQHC Comparison Sites (n=275,847)	PCC Comparison Sites (n=474,259)	FQHC Demonstration Sites (n=102,499)	FQHC Comparison Sites (n=163,334)	PCC Comparison Sites (n=290,365)
Diabetes	52,050 (34.2)	97,346 (35.3)	165,042 (34.8)	30,212 (29.5)	48,923 (30.0)	93,905 (32.3)
Moderate or end-stage liver disease	4,457 (2.9)	7,189 (2.6)	9,531 (2.0)	3,228 (3.1)	4,716 (2.9)	7,403 (2.5)
Neurological disorders	18,374 (12.1)	33,765 (12.2)	62,898 (13.3)	12,878 (12.6)	19,678 (12.0)	42,036 (14.5)
Cardiovascular disorders	20,282 (13.3)	38,875 (14.1)	84,971 (17.9)	11,808 (11.5)	18,910 (11.6)	52,125 (18.0)
Vascular disorders	16,160 (10.6)	29,612 (10.7)	72,343 (15.3)	9,443 (9.2)	14,724 (9.0)	44,595 (15.4)
Chronic heart failure	18,721 (12.3)	35,371 (12.8)	77,786 (16.4)	11,458 (11.2)	17,277 (10.6)	46,841 (16.1)
Trauma	8,061 (5.3)	14,737 (5.3)	32,162 (6.8)	5,451 (5.3)	8,314 (5.1)	22,526 (7.8)
Infections	4,275 (2.8)	7,773 (2.8)	19,388 (4.1)	2,927 (2.9)	4,174 (2.6)	13,037 (4.5)
Protein-calorie malnutrition	2,301 (1.5)	4,467 (1.6)	11,431 (2.4)	1,702 (1.7)	2,690 (1.6)	8,465 (2.9)
Renal failure	16,074 (10.6)	29,580 (10.7)	54,864 (11.6)	9,036 (8.8)	14,273 (8.7)	34,523 (11.9)
Gastrointestinal disorders	3,757 (2.5)	6,900 (2.5)	16,822 (3.5)	2,606 (2.5)	4,013 (2.5)	11,076 (3.8)
Pancreatic disease	2,151 (1.4)	3,911 (1.4)	7,748 (1.6)	1,512 (1.5)	2,374 (1.5)	5,532 (1.9)
Decubitus ulcer	4,009 (2.6)	7,915 (2.9)	18,866 (4.0)	2,646 (2.6)	4,079 (2.5)	12,484 (4.3)
Bone/joint/muscle infections or necrosis	1,325 (0.9)	2,446 (0.9)	5,138 (1.1)	1,021 (1.0)	1,622 (1.0)	4,066 (1.4)
Stroke	6,727 (4.4)	12,239 (4.4)	25,394 (5.4)	4,795 (4.7)	6,550 (4.0)	16,225 (5.6)
Hierarchical Condition Category score, mean (SD)	1.2 (1.0)	1.2 (1.0)	1.3 (1.2)	1.1 (1.0)	1.1 (1.0)	1.3 (1.2)

Exhibit B.3: Characteristics of Medicaid Claims-Based Beneficiary Cohorts Based on Site of First Attribution*

Characteristics	FQHC Demonstration Sites (n= 30,579)	FQHC Comparison Sites (n=69,704)
Age: 18–29 years, n (%)	7,115 (23.27)	19,304 (27.69)
30–39 years	6,803 (22.25)	15,889 (22.79)
40–49 years	7,020 (22.96)	14,484 (20.78)
50–59 years	5,453 (17.83)	12,370 (17.75)
60+ years	4,188 (13.70)	7,657 (10.99)
Gender: Male, n (%)	8,898 (29.10)	19,306 (27.70)
Female	21,681 (70.90)	50,398 (72.30)
Race/ethnicity: White, n (%)	8,644 (28.27)	21,715 (31.15)
Black	2,509 (8.20)	7,245 (10.39)
Asian	3,392 (11.09)	2,390 (3.43)
Hispanic	11,903 (38.93)	31,038 (44.53)
Pacific Islander	1,552 (5.08)	3,055 (4.38)
American Indian	221 (0.72)	441 (0.63)
Missing	2,358 (7.71)	3,820 (5.48)
Disabled, n (%)	9,540 (31.20)	22,867 (32.81)
Program: Fee-for-service, n(%)	6,435 (21.04)	15,423 (22.13)
Managed care	24,144 (78.96)	54,281 (77.87)
Any private health insurance, n (%)	582 (1.90)	1,865 (2.68)
Receipt Temporary Assistance for Needy Families, n(%)	4,936 (16.14)	13,066 (18.74)
Clinical conditions: HIV/AIDS, n (%)	147 (0.48)	582 (0.83)
Autoimmune disorders	623 (2.04)	1,385 (1.99)
Severe hematological disorders	89 (0.29)	239 (0.34)
Chronic lung disorders	1,428 (4.67)	3,377 (4.84)
Cancer	862 (2.82)	1,809 (2.60)
Chronic alcohol/drug dependence	1,353 (4.42)	2,922 (4.19)
Chronic mental health conditions	4,151 (13.57)	8,637 (12.39)
Diabetes	5,571 (18.22)	11,325 (16.25)
Moderate or end-stage liver disease	929 (3.04)	1,757 (2.52)
Neurological disorders	1,908 (6.24)	4,321 (6.20)
Cardiovascular disorders	989 (3.23)	2,129 (3.05)
Vascular disorders	693 (2.27)	1,559 (2.24)
Chronic heart failure	1,084 (3.54)	2,650 (3.80)
Trauma	686 (2.24)	1,634 (2.34)
Infections	447 (1.46)	972 (1.39)
Protein-calorie malnutrition	91 (0.30)	286 (0.41)
Renal failure	746 (2.44)	1,688 (2.42)
Gastrointestinal disorders	422 (1.38)	904 (1.30)
Pancreatic disease	279 (0.91)	667 (0.96)
Decubiti's ulcer	310 (1.01)	604 (0.87)
Bone/joint/muscle infections or necrosis	152 (0.50)	330 (0.47)
Stroke	664 (2.17)	1,314 (1.89)

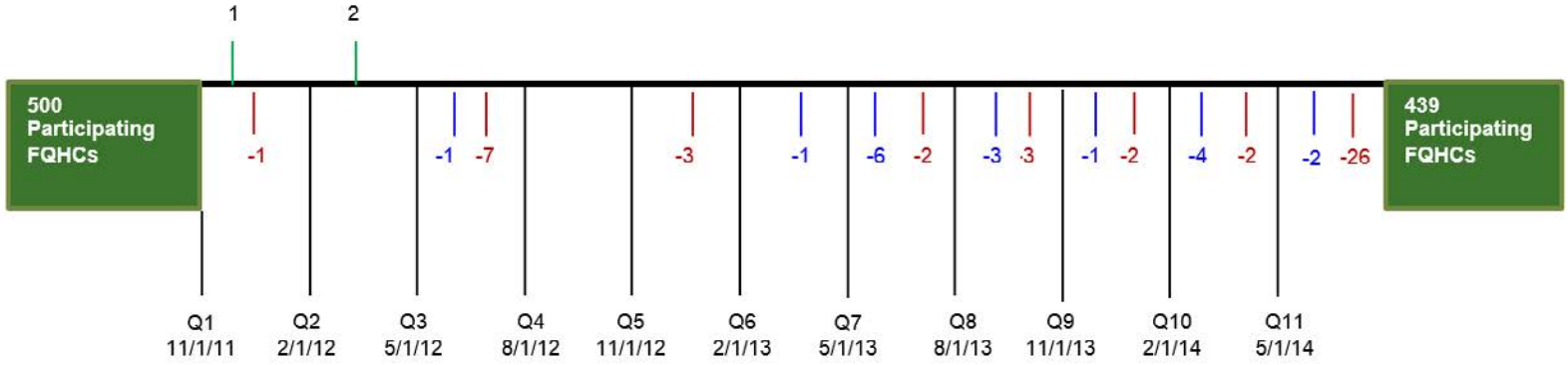
NOTE: ITT cohort from the state of California only. Cancer comorbidity excludes precancer or in situ status. These data reflect 56 demonstration sites and 92 comparison sites we were able to identify in the MSIS files.

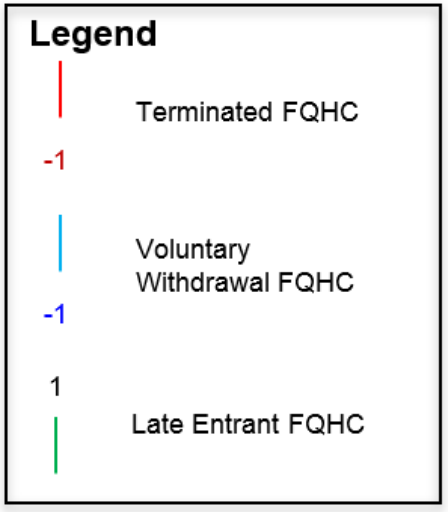
Exhibit B.4: Characteristics of Sites and Practices Included in the Medicare Beneficiary Cohort Analyses

Characteristic	FQHC Demonstration Sites (n=503)	FQHC Comparison Sites (n=827)	PCC Comparison Sites (n=1,083)
Location: Urban, n (%)	177 (35.2)	304 (36.8)	390 (36.0)
Rural	326 (64.8)	523 (63.2)	693 (64.0)
PCA region: Central, n (%)	127 (25.2)	167 (20.2)	304 (28.1)
Mid-Atlantic	61 (12.1)	110 (13.3)	112 (10.3)
Northeast	65 (12.9)	74 (8.9)	155 (14.3)
Southeast	76 (15.1)	181 (21.9)	256 (23.6)
West	85 (16.9)	129 (15.6)	136 (12.6)
West-Central	89 (17.7)	166 (20.1)	120 (11.1)
Racial composition of census tract: White, Mean % (SD %)	69.6 (28.6)	67.2 (28.0)	72.7 (26.6)
Black	16.5 (26.4)	18.1 (25.8)	14.5 (23.2)
Asian	3.1 (8.3)	3.3 (9.0)	3.8 (9.9)
American Indian	1.6 (6.8)	1.7 (7.0)	1.4 (7.0)
Hispanic	17.8 (25.3)	18.2 (25.5)	14.1 (23.3)
Foreign-born percentage in census tract, Mean % (SD %)	10.2 (13.3)	11.0 (13.9)	10.7 (15.4)
Household poverty in census tract, Mean % (SD %)	20.8 (11.3)	21.7 (11.5)	19.1 (11.6)
FQHC age: <5 years, n (%)	70 (13.9)	133 (16.1)	—
Age 5–10 years	125 (24.9)	198 (23.9)	—
Age 11–20 years	122 (24.3)	181 (21.9)	—
Age 21–30 years	61 (12.1)	107 (12.9)	—
Age 31–40 years	91 (18.1)	146 (17.7)	—
Age 41+ years	21 (4.2)	44 (5.3)	—
Missing age	13 (2.6)	18 (2.2)	—
Number of service delivery sites: 1 site, n (%)	11 (2.2)	62 (7.5)	—
2–5 sites	126 (25.0)	276 (33.4)	—
6–10 sites	148 (29.4)	270 (32.6)	—
11–20 sites	124 (24.7)	157 (19.0)	—
21+ sites	94 (18.7)	62 (7.5)	—
Number of providers: Primary Care, mean (SD)	5.5 (4.9)	5.4 (5.6)	3.8 (4.6)
Specialists	0.8 (1.8)	0.8 (2.2)	1.5 (3.4)
Midlevel	2.3 (3.0)	2.5 (2.9)	1.6 (2.6)
Behavioral Health/Social Service	0.3 (0.8)	0.3 (1.0)	0.0 (0.0)
Dental	0.0 (0.2)	0.0 (0.2)	0.0 (0.0)
Vision	0.1 (0.4)	0.1 (0.5)	0.0 (0.3)
Podiatry	0.2 (0.5)	0.1 (0.5)	0.1 (0.6)
Other	0.4 (1.0)	0.3 (0.9)	0.3 (0.8)
Total Patients per site, Mean (SD)	3,218.6 (2,204.7)	3,267.3 (2,289.6)	—
Medicare patients per site, Mean (SD)	277.8 (191.7)	292.9 (236.6)	—
Grant revenue per site in millions, Mean (SD)	0.8 (0.7)	0.8 (0.8)	—
Patient revenue per site in millions, Mean (SD)	1.3 (1.1)	1.3 (1.2)	—
FQHC Look-alike, n (%)	15 (3.0)	42 (5.1)	—
Safety Net Medical Home Initiative participant, n (%)	12 (2.4)	9 (1.1)	—
Beacon supplemental funding, n (%)	46 (9.1)	73 (8.8)	—

Characteristic	FQHC Demonstration Sites (n=503)		FQHC Comparison Sites (n=827)		PCC Comparison Sites (n=1,083)	
PCMH supplemental funding FY11, n (%)	471	(93.6)	560	(67.7)	—	
PCMH supplemental funding FY12, n (%)	383	(76.1)	547	(66.1)	—	
ACA-Building Capacity grantee, n (%)	100	(19.9)	117	(14.1)	—	
ACA-Immediate Facility Improvement grantee, n (%)	182	(36.2)	144	(17.4)	—	
ACA-New Access Point grantee, n (%)	71	(14.1)	122	(14.8)	—	
American Recovery and Reinvestment Act grantee, n (%)	320	(63.6)	587	(71.0)	—	
NCQA recognition (2008 standards): No recognition, n (%)	466	(92.6)	795	(96.1)	1,068	(98.6)
Level 1 recognition	5	(1.0)	5	(0.6)	4	(0.4)
Level 2 recognition	3	(0.6)	5	(0.6)	0	(0.0)
Level 3 recognition	29	(5.8)	22	(2.7)	11	(1.0)

Exhibit B.5: Demonstration Participation Over Time (as of September 29, 2014)





Source: RAND attrition tracking

Demonstration Quarter	Number of Voluntary Withdrawals	Number of Terminations
1	0	1
2	0	0
3	1	7
4	0	0
5	0	3
6	1	0
7	6	2
8	3	3
9	1	2
10	4	2
11	2	26
Total	**18**	**46**

Exhibit B.6: Characteristics of Sites and Practices Included in the Medicaid Beneficiary Cohort Analyses

Characteristic	FQHC Demonstration Sites (n=56)		FQHC Comparison Sites (n=92)	
Location: Urban, n (%)	27	(48.2)	53	(57.6)
Rural	29	(51.8)	39	(42.4)
Racial composition of census tract: White, Mean % (SD %)	62.8	(22.9)	64.7	(19.0)
Black	4.9	(8.1)	4.8	(9.1)
Asian	10.6	(18.3)	7.8	(9.5)
American Indian	2.1	(6.6)	1.6	(2.3)
Hispanic	40.6	(30.4)	47.5	(27.6)
Foreign-born percentage in census tract, Mean % (SD %)	27.4	(19.2)	29.4	(15.0)
Household poverty in census tract, Mean % (SD %)	20.0	(11.8)	19.1	(10.6)
FQHC age: <5 years, n (%)	6	(10.7)	8	(8.7)
Age 5–10 years	16	(28.6)	22	(23.9)
Age 11–20 years	11	(19.6)	18	(19.6)
Age 21–30 years	10	(17.9)	16	(17.4)
Age 31–40 years	9	(16.1)	19	(20.7)
Age 41+ years	2	(3.6)	9	(9.8)
Missing age	2	(3.6)	—	
Number of service delivery sites: 1 site, n (%)	1	(1.8)	5	(5.4)
2–5 sites	11	(19.6)	27	(29.3)
6–10 sites	17	(30.4)	24	(26.1)
11–20 sites	9	(16.1)	27	(29.3)
21+ sites	18	(32.1)	9	(9.8)
Number of providers: Primary care, mean (SD)	6.9	(5.3)	7.4	(5.3)
Specialists	1.1	(1.9)	1.0	(1.8)
Midlevel	2.6	(2.7)	2.2	(2.7)
Behavioral Health/Social Service	0.3	(1.0)	0.5	(0.9)
Dental	0.1	(0.3)	0.0	(0.1)
Vision	0.3	(0.7)	0.2	(1.0)
Podiatry	0.3	(0.5)	0.2	(0.4)
Other	0.9	(1.3)	0.3	(0.8)
Total patients per site, Mean (SD)	3,868.6	(1,667.3)	5,005.5	(2,927.9)
Medicare patients per site, Mean (SD)	284.6	(238.9)	284.9	(204.0)
Grant revenue per site in millions, Mean (SD)	0.9	(0.6)	1.0	(1.2)
Patient revenue per site in millions, Mean (SD)	1.8	(1.2)	2.2	(1.4)
FQHC Look-alike, n (%)	3	(5.4)	13	(14.1)
Safety Net Medical Home Initiative participant, n (%)	—		—	
Beacon supplemental funding, n (%)	12	(21.4)	7	(7.6)
PCMH supplemental funding FY11, n (%)	53	(94.6)	64	(69.6)
PCMH supplemental funding FY12, n (%)	47	(83.9)	65	(70.7)
ACA-Building Capacity grantee, n (%)	16	(28.6)	18	(19.6)
ACA-Immediate Facility Improvement grantee, n (%)	26	(46.4)	28	(30.4)
ACA-New Access Point grantee, n (%)	3	(5.4)	14	(15.2)
American Recovery and Reinvestment Act grantee, n (%)	25	(44.6)	64	(69.6)
NCQA recognition (2008 standards): No recognition, n (%)	56	(100)	91	(98.9)
Level 1 recognition	—		—	
Level 2 recognition	—		—	
Level 3 recognition	—		1	(1.1)

Regression Methodology

Regression analysis was used to examine whether there were statistically significant differences in beneficiary-level cost and utilization outcomes between comparison and demonstration FQHCs after controlling for covariates. Models used beneficiary cost and utilization outcomes as dependent variables and beneficiary, site, grantee, and area characteristics as covariates. Longitudinal models were used with repeated quarterly observations for each beneficiary.[32]

We modeled the impact of the demonstration on outcomes using a difference-in-differences model where the difference between the demonstration and comparison groups for each outcome is assumed to be constant during the baseline period, and is allowed to vary quarter-by-quarter in the demonstration period. This model is defined as:

$$Y = \alpha_0 + \alpha_1 I + \alpha_2 Q_2 + \alpha_3 Q_3 + \alpha_4 Q_4 + \alpha_5 Q_5 + \alpha_6 Q_6 + \alpha_7 Q_7 + \alpha_8 Q_8 + \alpha_9 Q_9 + \alpha_{10} Q_{10} + \alpha_{11} Q_{11} + \alpha_{12} Q_{12} + \alpha_{13} Q_{13} + \alpha_{14}(I * Q_5) + \alpha_{15}(I * Q_6) + \alpha_{16}(I * Q_7) + \alpha_{17}(I * Q_8) + \alpha_{18}(I * Q_9) + \alpha_{19}(I * Q_{10}) + \alpha_{20}(I * Q_{11}) + \alpha_{21}(I * Q_{12}) + \alpha_{22}(I * Q_{13}) + \gamma X + \epsilon \quad (1)$$

where

- *Y* is the beneficiary cost or utilization outcome
- α_0 is the intercept, an estimate of the mean adjusted level of *Y* in the comparison group in the first baseline quarter
- *I* is an indicator for the intervention, defined here as attribution to a demonstration FQHC (=0,1). Its parameter estimate, α_1, is an estimate of the difference in levels of cost/utilization associated with the demonstration group relative to the comparison group in the baseline period
- Q_t for time periods *t=2–13* is a binary indicator variable of the quarter of observation. For example, Q_2=1 for the second quarter and 0 for all other quarters. Parameters α_2 through α_{13} are estimates of the difference in beneficiary cost/utilization between the quarter of the indicator Q_t and quarter 1, in the comparison group
- $I*Q_t$ for time periods *t=5–13* is an interaction term that permits the impact of the demonstration to differ for demonstration sites in quarters 5–13, compared with the baseline period. **Parameters α_{14} through α_{22} are the estimates of interest in this model.** These parameters convey the impact of the demonstration on a quarter-by-quarter

[32] RAND's quarterly regression model is designed to provide rapid cycle feedback to the Center for Medicare and Medicaid Innovation on the demonstration's impact. Ultimately, we plan to use a model that estimates the impact of the demonstration *in the most recent year of the demonstration* relative to the baseline year. This approach has the advantage of smoothing over the quarterly measurements, which are measured less precisely.

basis in the demonstration period in relation to the baseline period. For example, $\alpha_{14}$is an estimate of how the difference between the demonstration and comparison groups in quarter 5 differs from the difference between the demonstration and comparison groups in the baseline period. α_{15} is an estimate of how the difference between the demonstration and comparison groups in quarter 6 differs from the difference between the demonstration and comparison groups in the baseline period.

- *X* is a vector of covariates. Its parameter estimates γ represent the difference in beneficiary cost/utilization associated with a one unit change in *X*;
- ϵ is a random error term that is assumed to follow an auto-regressive process where the error in one quarter is correlated with the error at the next quarter. The auto-correlation is estimated through the model.

This model outlined in Equation 1 allows the impact of the demonstration to vary in a nonlinear fashion from quarter to quarter in the demonstration period. It makes use of multiple quarters of a baseline period in which both the demonstration and comparison sites were observed without exposure to the intervention, as well as multiple quarters of an intervention period in which only demonstration sites are exposed to the intervention.

We used additional analytic techniques to further minimize the potential for bias caused by differences in characteristics between the demonstration and comparison groups. First, the model adjusts for differences in beneficiary, site, geographic, and other observed characteristics between demonstration and comparison FQHCs directly through vector X. Second, propensity score weights were used in conjunction with Equation 1 to differentially weight observations in the comparison group so that the mean characteristics of demonstration and comparison FQHCs and their attributed beneficiaries were comparable. Propensity scores were derived for each beneficiary using a logistic regression model that predicted participation in the demonstration as a function of beneficiary, site, grantee, and area characteristics:

$$(1)\ \ p = \Pr(I = 1) = \frac{1}{1+\exp(-\beta_0-\beta_1 X-\sum_{k=1}^{K}\beta_2 Y_k)}$$

where Y_k, $k=1, 2, ..., K$ are beneficiary outcomes in the baseline period and the vector of covariates X are listed in Exhibit B.7.

Exhibit B.7: Covariates Included in Propensity Score Models by Cohort

Beneficiary characteristics	Beneficiaries Attributed to FQHC Demo Sites and FQHC Comparison Sites	Beneficiaries Attributed to FQHC Demo Sites and PCC Comparison Sites
Age	X	X
Race	X	X
Gender	X	X
Dual eligibility status	X	X
Disabled	X	X
Institutionalization status	X	X
Number of primary care visits in prior year	X	X
Comorbidity index	X	X
In diabetes denominator	X	X
In ischemic vascular disease denominator	X	X
Total payments (4 quarterly measurements)	X	X
Number of inpatient admissions (4 quarterly measurements)	X	X
Number of emergency room (ER) visits (4 quarterly measurements)	X	X
Number of ambulatory care sensitive conditions (ACSC) admissions (4 quarterly measurements)	X	X
Number of readmissions (4 quarterly measurements)	X	X
Hemoglobin A1c test (4 quarterly measurements)	X	X
Nephropathy test (4 quarterly measurements)	X	X
Eye exam (4 quarterly measurements)	X	X
Low-density Lipoprotein test—diabetes cohort (4 quarterly measurements)	X	X
Low-density Lipoprotein test—IVD cohort (4 quarterly measurements)	X	X

Site characteristics	Beneficiaries Attributed to FQHC Demo Sites and FQHC Comparison Sites	Beneficiaries Attributed to FQHC Demo Sites and PCC Comparison Sites
Level 3 NCQA PCMH recognition at baseline (2008 standards)	X	
Number of beneficiaries per site (2010)	X	X
Total revenue per site	X	
Years FQHC has been operating	X	
Number of primary care physicians per site	X	X
Number of specialists per site	X	X
Number of nurse practitioners/physician assistants per site	X	X
Number of behavioral health/social service providers per site	X	
Number of dental providers per site	X	
Number of vision providers per site	X	
Number of podiatry providers per site	X	
HRSA PCMH Initiative participant	X	

Grantee characteristics	Beneficiaries Attributed to FQHC Demo Sites and FQHC Comparison Sites	Beneficiaries Attributed to FQHC Demo Sites and PCC Comparison Sites
Ambulatory Quality Accreditation	X	
Participation in CMS sharing savings demonstration	X	X
Number of service delivery sites	X	
HRSA Health Center Controlled Network grantee	X	

Area characteristics	Beneficiaries Attributed to FQHC Demo Sites and FQHC Comparison Sites	Beneficiaries Attributed to FQHC Demo Sites and PCC Comparison Sites
Rural-Urban Continuum Code	X	X
PCA Region	X	X
Percent household poverty in census tract	X	X
Medicaid payments to PCMHs in site	X	X

The propensity scores were derived from the fitted values of the regression model in Equation 2 and used as beneficiary-level weights in Equation 1. This "doubly robust" method provides unbiased estimates if the *propensity score model* fully captures the selection biases in the data. Additionally, a key advantage is that even if such an assumption is incorrect, estimates would remain unbiased as long as the *difference-in-differences model* in Equation 1 fully captures the impact of the demonstration. Our difference-in-differences model controls for potential differences in baseline outcomes between the demonstration and comparison groups, with model variable I. Model coefficient α_1 indicates the strength and significance of these baseline differences. The specification of all propensity score models was assessed through a combination of model fit statistics and post estimation assessments of balance. Imbalance was determined using absolute standardized differences. This approach emphasizes the importance of "practical differences" in covariates between demonstration and comparison groups rather than statistically significant differences. Most of RAND's use sample sizes that are extremely large and small difference are likely to produce positive significance tests. The standardized mean difference estimators used in our evaluation are:

$$Std\ Diff = \left| \frac{\bar{X}_D - \bar{X}_C}{SD(X)_{combined}} \right| \ for\ continuous\ variables$$

$$Std\ Diff = \ |p_D - p_C| \ for\ categorical\ variables$$

$$\bar{X}_D \ and \ \bar{X}_C \ are\ mean\ values\ for\ continuous\ variables \\ in\ the\ demonstration\ and\ comparison\ groups, respectively$$

p_D and p_C are mean proportions for categorical variables in the demonstration and comparison groups, respectively. In essence, for continuous variables, this standardized difference is computed as the difference in mean values for the variable between demonstration and comparison groups divided by the standard deviation of the combined sample, allowing all the

differences to be on the same scale. Any absolute standardized difference that is larger than 0.02, or 2 percent (i.e., the nonstandardized difference is larger than 0.02 times the standard deviations in absolute value), is assumed to be large in terms of its practical significance. For categorical variables, since these are already proportions bounded between
0 and 100 percent, the raw difference in proportions is computed and any difference larger than 2 percentage points in absolute value is also assumed to be practically significant. Essentially, a difference in proportion around 1 percentage point or less will be assumed small and any difference more than 2 percentage points was considered large and flagged for further review.

After controlling for these baseline differences, the effect of the intervention is estimated as the difference between the demonstration and comparison groups in each quarter of the demonstration period, compared to the difference between the demonstration and comparison groups in the baseline period. As noted above, these incremental changes for Quarters 5–13 are indicated through parameters α_{14} through α_{22} in the models.

The regression model described in Equation 1 was estimated using Generalized Estimating Equations (GEE) extension of the generalized linear model (GLM), with the family and link function varying by dependent variable. The family of the GLM specifies the distribution of the outcome variable, while the link function specifies the relationship between the mean of this distribution and the linear combination of predictor variables. Binary outcome data (readmissions) are modeled with a GLM model that uses a binomial distribution with a logit link function. Hospital admissions and emergency department (ED) visits are modeled using the negative binomial distribution with a log link, which is appropriate for right-skewed count data such as these. And finally, our cost data were modeled using a normal distribution with identity link. Model form specifications, including the family and link function used for each model, are summarized in Exhibit B.8. The specifications in the GEE model account for the autocorrelation structure in the errors due to repeated quarterly observations per beneficiary.

The models also included adjustments for changes in a beneficiary's eligibility for Medicare Parts A and B (and therefore the extent to which we observe the beneficiary's utilization and costs of health care services) within each quarter. The method of eligibility adjustment varied by model, as described in Exhibit B.8.

Exhibit B.8: Dependent Variables and Family, Link Function, and Eligibility Adjustment Used in Regression Model Specifications

Y Variable	Y Variable Type	Family	Link	Eligibility Adjustment
Medicare payments (total and individual categories)*	Continuous, skewed right	Normal	Identity	Divide Y by eligibility weight for beneficiaries that lose eligibility but remain alive during a quarter. No adjustment for beneficiaries that die during a quarter.
Admissions (all and chronic ACSCs)	Count	Negative binomial	Log	Number of months of eligibility included as offset
ER visits	Count	Negative binomial	Log	Number of months of eligibility included as offset

Y Variable	Y Variable Type	Family	Link	Eligibility Adjustment
Readmission within 30 days	Binary	Binomial	Logit	Not needed – measure requires eligibility during full 30-day observation period
Process measures	Binary	Binomial	Logit	Not needed – measure requires eligibility during one-year measurement period

*Payment categories include: acute care hospital, post–acute care, FQHC/ rural health center (RHC), outpatient, primary care physician, specialist physician

Parameter estimates from nonlinear GLM models are not always readily interpretable. For example, when modeling binary outcomes, such as readmission, with a binomial distribution and the logit link function, parameter estimates are expressed on the log-odds or odds ratio scales. To make the model estimates reliably comparable on the untransformed outcome scale, we used an estimator derived by Puhani for all nonlinear models (e.g., all-cause admissions, ambulatory care sensitive conditions [ACSC] admissions, ED visits, readmissions, process measures) [33] that provides an analog to a traditional difference-in-differences estimator but is appropriate for nonlinear models. Puhani's method uses two sets of predictions to estimate the treatment effect. First, the full nonlinear GLM model (Equation 1) is used to estimate the mean predicted value of the dependent variable, assuming that all beneficiaries were in the demonstration in the specific quarter of interest (for example, Quarter 5). The second prediction estimates a mean value of the dependent variable under a "counterfactual" scenario in which there was no demonstration—by setting the interaction term between the demonstration and the quarter of interest in the prediction model equal to zero, as suggested in Puhani. The difference between these two predictions will be the retransformed estimator of the parameter of interest (in this example, the interaction term between the demonstration and Quarter 5) in the original outcome scale. This procedure is repeated for each quarter in the demonstration period to obtain estimates of the incremental impact of the demonstration for each quarter in the demonstration period, relative to the baseline period. We implemented the Puhani estimator in Stata using the margins command, which produces mean estimates of each dependent variable (on the untransformed measurement scale) that average over all beneficiary-level predictions. Standard errors computed using this command are estimated using the delta method.

For linear models that modeled the demonstration's impact on total Medicare payments and payments within six categories, we used GLM models (without the use of the margins command). Standard errors for these estimates are obtained analytically and without approximation.

[33] Puhani, P. A. (2012). The treatment effect, the cross difference, and the interaction term in nonlinear 'difference-in-differences' models. *Economics Letters*, *115*, 85–87.

Appendix C. Baseline Beneficiary Survey

To collect information on patient experience of care at FQHCs, we conducted a patient survey using items from validated instruments and focusing on aspects of patient experience especially pertinent to FQHCs. Collecting information on the impact of the demonstration on patient experience of care is a critical component of the evaluation. Patients are the best source of this information. Patient-experience-of-care data collected with the expanded Clinician and Group Consumer Assessment of Healthcare Providers and Systems (CG-CAHPS) Survey is being used to evaluate whether FQHCs participating in the demonstration provide Medicare beneficiaries with

- more timely delivery of health services
- better coordination of care
- better experiences with the health care system, including more effective participation in decisions about health care.

We have completed fielding the baseline survey and are in the process of fielding the follow-up survey. This report analyzes data from the complete baseline survey, whereas the first annual report presented analyses from only a portion of the baseline survey cohort.

In fielding the sample of Medicare beneficiaries during the baseline survey, we followed the CAHPS guidelines for data collection. The survey was fielded concurrently in English and Spanish, using a mixed-mode data-collection approach (mail with telephone follow-up to nonrespondents).

Population Surveyed

We selected the beneficiary survey sample by selecting Medicare beneficiaries from practices attributed to demonstration and comparison sites, including FQHC and PCC comparison sites. As with the larger evaluation, inclusion of PCC sites provides the opportunity to include a comparison group of practices less likely to be exposed to TA comparable to that provided by the advanced primary care practice (APCP) demonstration, even though the PCC practice sites are likely to differ somewhat from FQHCs. To select the beneficiary survey sample, we first matched demonstration sites to comparison sites using propensity score methods. Then, within each site, we selected participants randomly (estimating a completion of 14 surveys per site) while stratifying on the characteristics we planned to oversample:

- age (i.e., less than 65 versus age 65 or older)
- dual Medicare eligibility (i.e., Medicare with Medicaid eligibility versus Medicare without Medicaid eligibility)

- hierarchical condition categories (HCC) scores (in the 75th percentile versus below the 75th percentile)
- probability of Spanish-language preference (high versus low).

To be eligible for the survey, beneficiaries had to have been attributed to either a demonstration FQHC intervention site or a comparison site according to the plurality rule.[34] For the baseline survey, we created a sample file of 28,235 beneficiaries attributed to demonstration or comparison FQHCs and listing first and last name, date of birth, age or disabled eligibility status, Medicare-only or dual eligibility, HCC scores, probability of Spanish-speaking preference, and mailing address. A similar file has been created for 2,412 beneficiaries attributed to comparison PCCs. We stratified our analyses according to beneficiary characteristics to have enough of a sample to conduct subgroup analyses in different groups.

The baseline beneficiary survey was fielded in two parts: the original cohort and the supplemental cohort. Both baseline beneficiary survey cohorts were fielded from the same main sample that included 30,647 Medicare beneficiaries, 28,235 of whom were attributed to the demonstration or comparison FQHC sites, and 2,412 of whom were attributed to PCCs.

Survey Respondent Characteristics Associated with the Full Baseline Beneficiary Survey Cohort Stratified by Demonstration FQHC, Comparison FQHC, and Comparison PCC

Exhibit C.1 shows beneficiary survey respondent characteristics for demonstration FQHC compared with comparison FQHC users. All percentages reported are weighted percentages, applying weights that account for survey stratification and nonresponse. We observe no significant difference in gender or education levels. There are also no differences in age between demonstration and comparison FQHCs.. The only racial differences are that comparison FQHC beneficiaries are less likely than demonstration FQHC beneficiaries to be American Indian or Alaska natives.

There are very few differences in comorbid characteristics between baseline FQHC and comparison FQHC beneficiaries. Among 33 reported comorbidities, comparison FQHC beneficiaries are less likely to have stomach ulcers and more likely to control their diabetes with medication compared to demonstration FQHC beneficiaries. Comparison PCC beneficiaries are more likely to have osteoporosis and less likely to have protein in their urine compared to demonstration FQHC beneficiaries. Once we account for multiple comparisons, these differences are not expected to be important.

---

[34] RAND's plurality rule assigns a beneficiary to the provider who offers the greatest number of primary care services over a 12-month period. RAND's attribution rule allows beneficiaries to be attributed to one of four types of providers: demonstration FQHCs or one of three types of comparison sites (FQHCs not participating in the demonstration, RHCs, or PCCs). By contrast, the attribution rule used by CMS (which ultimately determines the allocation of care management fees to demonstration FQHCs) restricts the sample of providers eligible for attribution to demonstration FQHCs alone.

There are very few significant differences in health status, mental health, and social stressors between demonstration FQHC and comparison FQHC or demonstration FQHC and comparison PCC beneficiaries, although comparison PCC beneficiaries do report having fewer social stressors than demonstration FQHC beneficiaries.

Exhibit C.1: Survey Respondent Characteristics Associated with the Full Baseline Beneficiary Survey Cohort Stratified by Demonstration FQHC, Comparison FQHC and Comparison PCC

Cohort Responding to the Survey[1] Demographics	Demonstration FQHC (N =7946) %	Comparison FQHC (N =8116) %	p-value[2] (Demonstration vs. Comparison FQHC)
Male (%[3])	39.02	39.17	0.9137
Female (%)	60.98	60.83	
Age 18-24 (%)	0.23	0.21	0.0789
Age 25-34 (%)	2.35	1.91	
Age 35-44 (%)	4.47	5.01	
Age 45-54 (%)	12.80	10.60	
Age 55-64 (%)	18.36	18.52	
Age 65-75 (%)	36.09	35.29	
Age 75 or older (%)	25.70	28.46	
8th grade or less (%)	15.26	16.57	0.5759
Some high school, but did not graduate (%)	16.99	16.96	
High school graduate or GED (%)	35.44	35.39	
Some college or 2-year degree (%)	22.21	21.83	
4-year college graduate (%)	4.42	4.68	
More than 4-year college degree (%)	5.67	4.57	
Hispanic (%)	16.49	15.35	0.5644
White (%)	71.61	70.55	0.6297
Black or African American (%)	14.17	15.56	0.4804
Asian (%)	2.03	1.61	0.5128
Native Hawaiian or Other Pacific Islander (%)	0.36	0.61	0.1845
American Indian or Alaskan Native (%)	5.17	3.94	0.0361
Other (%)	6.43	5.74	0.3513

Cohort Responding to the Survey[1] Comorbidity	Demonstration FQHC (N =7946) %	Comparison FQHC (N =8116) %	p-value[2] (Demonstration vs. Comparison FQHC)
Any comorbidity (%)	96.07	95.64	0.4545
Any heart comorbidity (%)	95.44	94.66	0.2291
Any kidney comorbidity (%)	67.99	68.92	0.6582
Any lung comorbidity (%)	75.64	74.21	0.4579
Any gut comorbidity (%)	65.32	61.08	0.0615
Any brain comorbidity (%)	75.72	74.35	0.4786
Any bone comorbidity (%)	94.08	93.55	0.4718
Any other comorbidity (%)	93.81	93.90	0.9095
Any diabetes (%)	35.86	36.37	0.7142
Myocardial infarction (%)	13.12	14.08	0.3751
Congestive heart failure (%)	12.33	11.59	0.4664

Cohort Responding to the Survey[1] Comorbidity	Demonstration FQHC (N =7946) %	Comparison FQHC (N =8116) %	p-value[2] (Demonstration vs. Comparison FQHC)
Angina (%)	13.50	13.00	0.6397
Coronary artery disease (%)	18.74	18.29	0.7126
Peripheral vascular disease (%)	14.81	14.96	0.8867
Hypertension (%)	83.99	84.44	0.6917
Kidney problems (%)	48.41	49.46	0.6427
Protein in your urine (%)	28.87	27.73	0.5658
Chronic lung disease (%)	38.67	38.15	0.8263
Asthma (%)	51.86	50.08	0.4192
Stomach ulcers (%)	47.86	42.44	0.0223
Liver problems (%)	23.50	23.21	0.8913
Headaches (%)	32.38	31.88	0.8014
Seizures (%)	12.02	11.07	0.4952
Dementia (%)	7.49	8.62	0.2842
Learning disability (%)	16.50	16.76	0.8880
Osteoporosis (%)	28.64	28.67	0.9802
Back problems (%)	47.79	47.62	0.9187
Arthritis (%)	71.46	71.14	0.8193
Thyroid disease (%)	24.49	24.63	0.9286
Anemia (%)	19.19	18.39	0.5116
Eye problems (%)	49.39	49.68	0.8660
Difficulty hearing (%)	35.55	35.29	0.8774
Another condition (%)	21.23	21.47	0.8618
Diabetes (%)	35.86	36.37	0.7144
Control diabetes with diet and exercise (%)	62.82	60.24	0.2400
Control diabetes with medication (%)	69.12	73.18	0.0389
Control diabetes with insulin (%)	33.07	30.70	0.2549
Calculated body mass index (BMI) (mean)	29.76	29.80	0.8517
Calculated BMI—neither overweight or obese (mean)	24.96	26.27	0.5540
Calculated BMI—overweight (mean)	31.99	32.25	
Calculated BMI—obese (mean)	43.05	41.48	

Cohort Responding to the Survey[1] Health Status	Demonstration FQHC (N =7946) %	Comparison FQHC (N =8116) %	p-value[2] (Demonstration vs. Comparison FQHC)
12-Item Short-Form Health Survey (SF-12) Physical Health (%)	32.60	32.68	0.8514
SF-12 Mental Health (%)	36.25	35.96	0.4531
Overall health excellent (%)	5.32	4.30	0.0526
Overall health very good (%)	14.61	16.03	
Overall health good (%)	34.86	32.02	
Overall health fair (%)	33.94	35.87	
Overall health poor (%)	11.26	11.77	
Overall mental health excellent (%)	14.77	15.16	0.9248

Cohort Responding to the Survey[1] Health Status	Demonstration FQHC (N =7946) %	Comparison FQHC (N =8116) %	p-value[2] (Demonstration vs. Comparison FQHC)
Overall mental health very good (%)	23.26	23.81	
Overall mental health good (%)	31.59	30.81	
Overall mental health fair (%)	23.97	24.18	
Overall mental health poor (%)	6.40	6.03	
Your physical health now limits you in moderate activities (moving a table, pushing a vacuum cleaner, etc.) (%)	69.43	69.47	0.8338
Your physical health limits you in climbing several flights of stairs (%)	72.08	73.57	0.4739
You accomplished less than you would like in the past 4 weeks because of your physical health (%)	63.12	63.38	0.8542
You were limited in the kind of work or other activities in the past 4 weeks by your physical health (%)	65.38	64.63	0.5790
You accomplished less than you would like in the past 4 weeks because of your mental health (%)	37.27	36.17	0.4123
Did work or other activities less carefully than usual in the past 4 weeks because of your mental health (%)	30.85	31.83	0.4665
Pain interfered with your normal work at least a little bit in the past 4 weeks (%)	81.20	80.48	0.6007
Felt calm and peaceful at least some of the time in the past 4 weeks (%)	83.69	82.92	0.6949
Had a lot of energy at least some of the time in the last 4 weeks (%)	66.02	63.79	0.3729
Felt downhearted or blue at least some of the time in the last 4 weeks (%)	43.51	43.13	0.2974
Physical and emotional problems interfered with your social activities at least some of the time in the past 4 weeks(%)	47.43	47.14	0.9753
Patient Health Questionnaire for Depression and Anxiety (PHQ-4) scale (Mean)	3.18	3.29	0.2951
Patient Health Questionnaire for Depression and Anxiety (PHQ)-anxiety (Mean)	1.64	1.70	0.2790
PHQ-depression (Mean)	1.57	1.63	0.2903
Felt nervous, anxious or on edge more than half the days in the last 2 weeks (%)	21.88	23.56	0.4146
Not able to stop or control worrying more than half the days in the last 2 weeks (%)	23.78	25.81	0.2000
Had little interest or pleasure in doing things more than half the days in the last 2 week (%)	24.82	25.92	0.6743

Cohort Responding to the Survey[1] Health Status	Demonstration FQHC (N =7946) %	Comparison FQHC (N =8116) %	p-value[2] (Demonstration vs. Comparison FQHC)
Felt down, depressed or hopeless more than half the days in the last 2 weeks (%)	19.66	21.51	0.0404
Money problems in the past month (%)	42.81	42.12	0.6532
Job problems in the past month (%)	6.16	5.92	0.7156
Problems with the police in the past month (%)	0.82	0.83	0.9538
Been the victim of a crime in the past month (%)	2.40	2.30	0.7883
Family or marriage problems in the past month (%)	12.13	10.28	0.0320
Victim of violence in your home in the past month (%)	1.45	1.59	0.6212
Witnessed violence in your home in the past month (%)	1.17	1.21	0.8713
Problems with your children in the past month (%)	9.47	8.22	0.0772
Problems with your grandchildren in the past month (%)	5.71	4.89	0.1724
Problems with someone else's children in your home in the pasta month (%)	1.88	2.01	0.7314
Mean Stress Scale of 10 items (SD)	0.84 (2.11)	0.79 (1.97)	0.4295
% with 0 of 10 points on stress scale	50.50	51.63	
% with 1 of 10 points on stress scale	29.66	30.30	
% with 2 of 10 points on stress scale	11.35	10.60	
% with 3 of 10 points on stress scale	4.94	4.14	
% with 4 of 10 points on stress scale	2.16	2.27	
% with 5+ of 10 points on stress scale	1.39	1.05	

In July 2014, we fielded the Beneficiary Follow-Up Survey. The sample for the follow-up survey consisted of all beneficiaries who completed the baseline survey, including 13,262 beneficiaries belonging to the original sample and 4,033 belonging to the supplemental sample.

In this report, baseline beneficiary survey data are collected from individuals attributed to the demonstration FQHC and comparison FQHC clinics. RAND's plurality rule for attribution assigns a beneficiary to the provider who offers the greatest number of primary care services over a 12-month period. The collection of 17,295 beneficiary survey responses spanned ten months, beginning in month 19 of the 36-month demonstration, and includes 7,948 responses from users of demonstration FQHCs, 8,117 responses from users of comparison FQHCs.. The collection of follow-up beneficiary survey responses will also span ten months beginning at the end of the demonstration (month 36 after demonstration initiation). The timing of the follow-up survey toward the end of the demonstration and beyond allows us to evaluate sustainable changes in the demonstration should they exist.

Baseline beneficiary survey analysis variables presented in this report are treated as dependent variables, allowing us to see whether differences between demonstration and comparison sites are recognizable. This report presents bivariate data comparing demonstration

FQHC with comparison FQHC users.. All percentages reported are weighted percentages, applying weights that account for survey stratification and nonresponse. P-values are based on robust standard error estimates that account for site-level clustering using chi-square tests for continuous variables and t-tests for dichotomous and continuous variables.

Since these analyses represent beneficiary reports midway through the demonstration, any differences between demonstration and comparison groups may represent an early demonstration effect. Alternatively, recognizing the slow start for systematic TA applied to demonstration sites, even if a demonstration effect may become apparent, these baseline data may be too early in the demonstration to recognize any difference between demonstration and comparison sites.

With time and the coming availability of the follow-up baseline survey, we will apply a difference-in-differences analysis to the longitudinal beneficiary survey data as we are currently implementing for claims data.

Appendix D. RAS/NCQA Data

Readiness Assessment Survey and Recognition

NCQA PCMH recognition is based on scoring according to six standards, each of which comprises multiple elements. Recognized sites achieve Level 1, 2, or 3 recognition based on their total number of points scored across elements and on the number of points scored on must-pass elements. Of 27 total elements, six are must-pass items considered essential to the functioning of PCMHs, and are required for practices at all recognition levels. Each of the six must-pass elements maps to a distinct PCMH standard. Practices must achieve a score of 50 percent or higher on must-pass elements. RAND receives reports of NCQA level recognition from demonstration FQHC, comparison FQHC, and comparison PCC sites.

As a condition of the demonstration, all participating FQHCs are required to complete the RAS every six months. Each site uses NCQA's web-based tool to complete the survey. "Preliminary" scores are then calculated that allow sites to anticipate the outcome of an application to NCQA for formal PCMH recognition. RAS scores are available for demonstration sites from the time of their initial application to become a demonstration site through May 2014. While NCQA recognition scores are available only as a single value that can range from 0 to 3, RAS scores generate six domain-specific values as well as an aggregate value.

In addition to NCQA recognition, several other organizations provide recognition awards to practice sites as a way of acknowledging achievement of competence in APCP attributes.

Sites are required by the demonstration to both complete the biannual RAS and to apply for NCQA recognition (with the expectation that they will attain Level 3 recognition). The most recent RAS deadline was May 1, 2014. Sites that applied for NCQA recognition during the three months prior to the RAS deadline (February 1, 2014, through April 30, 2014) were not required to submit a new RAS. For sites with missing RAS scores, Truven (the CMS contractor that collects the RAS data) fills missing RAS data scores with the most recent prior site-level RAS score. The May 2014 RAS is the final required survey for the demonstration. If sites submitted during this three-month window and received their recognition result, their recognition extract is substituted for their RAS score.

Anecdotally, we have learned from the PCAs and AIR that some FQHC sites are delaying their NCQA recognition. There are a number of reasons that sites would make this decision. First, they are funded for only one application, so some may wait and ensure they are as prepared as possible. Although sites can resubmit for higher levels of recognition without penalty or cost, some have noted they see little value in applying earlier and risking attainment of only Level 1 or Level 2 recognition. Finally, there is an expiration date on NCQA recognition, so by waiting until the end of the demonstration, sites can maximize their time before needing to reapply.

Appendix E. CASE Methods

Overview of CASE Survey

The CASE survey addresses multiple aspects of site-level transformation to APCPs but also documents specific aspects of demonstration site experiences with components of CMS's FQHC intervention.

CASE Survey Development

We used versions of the CASE survey instruments assembled by the PCMH Evaluators' Collaborative: Measurement Workgroup on Clinician and Staff Experience.[35] One version of the survey was intended for respondents who were physicians, nurse practitioners, or physician assistants, and the other version was for other clinic staff (including both clinical and nonclinical staff). Both were paper-based surveys primarily containing items with discrete choice response categories.

The baseline CASE survey included previously validated scales to measure the five domains:

- work content
- work perceptions, including satisfaction, stress, burnout, intent to leave, work control, chaos, time pressure
- culture, including values alignment with leaders, communication openness and organizational learning, care team functioning, and adaptive reserve
- demonstration-specific domains, including knowledge of participation in the APCP demonstration, use of demonstration feedback reports, attendance and use of TA
- FQHC-specific domains, including site efforts to seek or avoid complex patients, and to increase or decrease specific types of care (in-office, out-of-office, hospital, specialist), communication with hospitals, and access to specialists.

We repeated the survey items associated with the baseline CASE for the follow-up survey but added items to gather additional information from respondents about ongoing changes in the structure of demonstration FQHCs and the experiences of clinicians and staff who work there. New items for the follow-up CASE survey focus on barriers to receipt of specialty referrals, tests, and procedures, as well as perceived barriers to medical home transformation.

[35] Commonwealth Fund. (2011). The patient-centered medical home evaluators' collaborative, website. As of December 4, 2014: http://www.commonwealthfund.org/Publications/Other/2010/PCMH-Evaluators-Collaborative.aspx.

CASE Sample

We randomly selected six respondents (three clinicians, three staff) from each of the originally selected 500 participating FQHC sites. CASE surveys were fielded even on sites that dropped out of the demonstration. The only exclusions were those initially selected FQHCs that were replaced within 30 days of being selected; for these sites, we will include the replacement sites. Because some FQHC sites had fewer than six eligible respondents, we applied any "freed up" surveys (e.g., a site with only four eligible respondents would "free up" two surveys) to other sites, so that in some sites we fielded the survey to seven eligible respondents. We fielded the CASE survey from April to September 2013.

Across the 503 participating FQHCs, we surveyed 3,011 eligible respondents including 1,496 clinicians and 1,515 other staff. We received 1,327 responses (44 percent overall response rate) including receipt of surveys from 562 clinicians and 765 other staff. We received at least one completed survey from 442 unique demonstration sites.

In conducting follow-up CASE surveys, we will prioritize resources, obtain a completed clinician survey from each clinic in the follow-up sample first, and—as resources permit—turn our attention to obtaining a completed survey from other clinic staff. However, our overall goal is to obtain a survey response from as many clinics as possible and we will aim to get at least one clinician and one staff survey per clinic.

Analysis of CASE Variables

In this report, analytic variables derived from baseline CASE survey responses are treated as dependent variables. Although longitudinal CASE data are not yet available, cross-sectional associations between RAS levels and CASE responses may foreshadow how changes in practice toward "medical homeness" affect dimensions of clinician and staff experience.

In "unadjusted" models for each CASE-derived dependent variable, the independent variables are only those listed in the exhibit, taken one at a time (e.g., region, urbanicity, RAS level). In "adjusted" models for each CASE-derived dependent variable, the independent variables are survey version (physician/nurse practitioners/physician assistant survey or other staff survey), practice size, and the other variables listed in the exhibits (e.g., region, urbanicity, RAS level) included simultaneously. The adjusted models also apply weights to account for survey nonresponse. In all models, the functional forms are chosen to match the distribution of the dependent variable (e.g., logistic link to match dichotomous variables; identity link to match normally distributed continuous variables as in scale scores), and the unit of analysis is the individual survey respondent, employing GEE to adjust standard errors for site-level clustering.

RAS levels are from November 1, 2012 to May 1, 2013, the closest RAS measurement period antecedent to CASE "baseline" survey, which began fielding in May 2013.

Appendix F. Qualitative Methods

Qualitative Methods Overview

To better understand the process and implementation issues involved in PCMH transformation, NCQA recognition, and delivery of TA, the evaluation has utilized several qualitative methods tailored to the types of stakeholders and levels of the APCP demonstration. These methods include interviews, focus groups, and site visits.

Interviews: Semi structured interviews have been used to collect qualitative perspectives from site leaders (from both demonstration and comparison sites) and from leaders of state PCAs that are involved in delivering TA directly to sites. The interviews last approximately one hour each and are conducted by telephone, except for those conducted in person during a subset of site visits (see below).

Focus groups: Focus groups have been used to collect qualitative perspectives from practice coaches employed by state PCAs to deliver direct TA to demonstration sites, and are being used to understand perspectives of Medicare beneficiaries and caregivers receiving care from demonstration sites. The PCA practice coach focus groups are conducted by teleconference call supplemented with web-meeting slide prompts, and have included four to ten participants, depending on the size of the PCA cluster region. All beneficiary and caregiver focus groups are being conducted in person at their respective FQHC demonstration sites, and are designed to include eight to 12 participants per focus group. Both the PCA practice coach and beneficiary/caregiver focus groups were designed to last approximately 90 minutes.

Site visits: We also have been conducting full-day site visits at five demonstration sites that include in-person interviews, walk-through and other observations of clinic operations and process, and PCMH-related documents that sites are willing to share (e.g., clinic brochures, patient-provider agreements).

Nested sampling design: Per our initial sampling plan described in the Evaluation Design Report, the qualitative components of the evaluation utilized a nested sampling design, starting at the regional level and identifying a set of six states—one from each region—as an integrated sample frame for the PCA interviews, focus groups, and other data-collection activities (Exhibit F.1). At the site level, we purposely selected 20 sites from the participating FQHC sample and ten from the comparison FQHC sample using a trifold stratification method. We first selected the 20 intervention sites using the following three stratification criteria: *state* (the six states—one from each of the six PCA clusters—identified in the integrated sample frame); *RAS score trajectory* (three categories: Low Baseline–Low Year 1, High Baseline–High Year 1, and Change ≥ 15 RAS scores between baseline and year 1); and *urbanicity* (two categories: rural and urban, based on U.S. Census indicators for geocoded addresses of each FQHC). Low RAS scores

are defined as those within the bottom tertile for all demonstration sites, and high RAS scores as those within the top tertile.

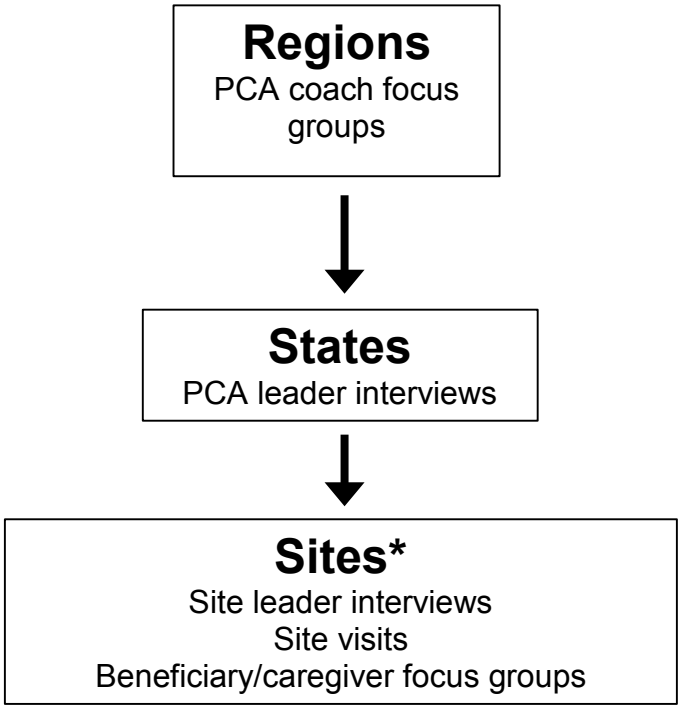
From Site to Grantee-Level Sampling: After several initial demonstration site leader interviews, we learned that, for the majority of sites, the individuals primarily responsible for PCMH change efforts and participation in the FQHC APCP demonstration were located at the FQHC *grantee level*, rather than being limited to the single site selected by RAND for the baseline interview. Thus, we moved from a site-level to an FQHC grantee–level sampling strategy. We examined the characteristics of the selected demonstration sites and determined that the original sample of sites provided a diverse sample along three grantee-level characteristics that RAND considered to be most important: (1) number of service delivery sites operated by the grantee, (2) percentage of sites located in rural areas, and (3) percentage of sites participating in the HRSA PCMH Initiative. Sampled demonstration grantees that declined to participate or failed to respond to our invitations (n=11) were replaced by other demonstration grantees whose profile matched the original according to these three characteristics. One sampled state did not include any demonstration.

Similarly, the ten comparison grantees we selected for baseline site interviews were those that most closely matched ten of the 20 demonstration FQHC grantee organizations. Sampled comparison FQHCs that declined to participate or failed to respond to our invitations were replaced by others that had the next closest match to ten of the 20 demonstration FQHC grantee organizations.

Recruitment: For each site selected, RAND contacted the designated FQHC APCP demonstration contact (or for comparison FQHCs, the main contact listed by HRSA) to identify the most appropriate leader familiar with PCMH change and/or practice transformation to invite for an interview. The main interviewee identified by this process frequently invited one or two other individuals within their organization also involved in PCMH change efforts. Interview respondents from demonstration FQHCs did not receive any payment for participating in the interview. Comparison FQHCs received an incentive payment of $25 per person for up to two persons per interview.

For each of the six states in the integrated sample frame, RAND and CMS worked with the lead TA contractor, AIR, to identify the most appropriate state PCA leaders. CMS sent a letter by email to each selected state PCA leader encouraging the PCA to participate in an interview with RAND. RAND then invited each identified state PCA leader individually to participate in an interview. The state PCA leaders were also allowed to include other PCA staff knowledgeable of their PCMH TA in the interview.

Exhibit F.1: Qualitative Nested Sampling Design



* Site data collection included a focus on the *grantee level*, since the main informants of PCMH efforts for many sites were at this level rather than the individual clinic locations initially selected.

Exhibit F.2 details the selection criteria and sample size for the qualitative methods used for each stakeholder group. Qualitative data were collected at two time points (early and late) for all stakeholder and method combinations, except for the site visits and the beneficiary/caregiver focus groups which are only being conducted at one (late) time point.

Exhibit F.2: Qualitative Sample Selection Criteria and Sizes

Stakeholders	Method	Selection Criteria	Sample Size
Regions			
PCA practice coaches	Focus groups	1 group per region including providers of direct TA	n=6 groups†
States			
PCA leaders	Interviews	1 state per region including 3 cluster lead PCAs and 3 other PCAs	n=6 PCAs†
Sites			
Demo site leaders	Interviews	State,* RAS score trajectory (high-high, low-low, improver), and urbanicity	n=20 sites†
Comparison site leaders	Interviews	Matched half selected demo sites on selection criteria	n=10 sites†
Demo site leaders and staff	Site visits	RAS Score trajectory (2 improvers, 2 high-high, 1 low-low); variation by state, size, urbanicity	n=5 sites‡
Beneficiaries/ caregivers	Focus groups	5 sites, 3 per site including2 beneficiary and one caregiver focus groups)	n=15 groups‡

* Sampled from the 6 selected states; however no demo sites were able to be recruited from one state.
† Two time points, early (year 2) and late (year 3)
‡ One late (year 3) time point

Exhibit F.3 provides details on the timing of each data collection.

Exhibit F.3 Timing of Qualitative Date Collection

Type of Qualitative Data Collection	Timing	Number and Type of Respondents
Late Site Leader Interviews	Oct-Dec 2014	20 leaders of demonstration sites; 10 leaders of comparison sites
Site visits	Aug-Nov 2014	5 demonstration sites
Early PCA Interviews	Aug-Oct 2013	6 PCA leaders (one in each of the 6 clusters)
Late PCA interviews	Sep-Nov 2014	6 PCA leaders (one in each of the 6 clusters)
Early PCA Focus Groups	Sep-Nov 2013	6 focus groups (one in each of the 6 clusters)
Late PCA Focus Groups	Oct-Dec 2014	6 focus groups (one in each of the 6 clusters)
Early Site Leader Interviews	May-Sep 2013	20 leaders of demonstration sites; 10 leaders of comparison sites
Beneficiary Focus Groups	Nov-Dec 2014	15 focus groups (three each at 5 demonstration sites)

The second annual report presents results on the qualitative data collection and analysis completed at the time the report was submitted, namely the thematic analysis of the early site leader and early PCA leader interviews. Analyses of the later site leader and later PCA leader interviews, along with other qualitative analyses, will be included in the final report.

FQHC Site Leader Interviews

As planned in RAND's Evaluation Design Report, we conducted a set of baseline site interviews with leaders responsible for PCMH implementation and/or practice transformation in 20 demonstration FQHCs and ten comparison FQHCs, between May and September 2013. The purpose of the interviews with demonstration sites was to understand the context, intervention, and implementation process of participating FQHCs with PCMH transformation and recognition. The purpose of the interviews with the comparison sites was to help identify potential unique drivers of change in intervention sites and to improve generalizability of demonstration site findings to the wider population of FQHCs beyond the APCP demonstration.

Topics for the baseline site interviews included an overview of clinic context and characteristics, such as engagement with other initiatives (organized by CMS or other bodies); perspectives on the PCMH model, recognition, and effects on patient and family experiences of care; practice change and PCMH implementation issues; and (for demonstration sites) experiences with the demonstration, including use and perceived utility of the TA, enhanced PCMH payments and quarterly feedback reports, and the effect of the FQHC APCP demonstration on clinic operations.

Analysis of Site Leader Interviews

We used a variation of content analysis to develop a coding scheme for performing a qualitative description of the themes discussed by the FQHC leaders. In this approach, we first developed an initial codebook based on the items in the interview protocol. Three evaluation team members, led by Dr. Peter Mendel, independently test coded the same two transcripts (conducted by separate interviewers) for all major themes in the codebook using the NVivo qualitative software package. Inter-rater reliability across all codes ranged from 72 percent to 89 percent agreement. Discrepancies were resolved by consensus in discussion among the three coders, which also resulted in additions or modifications to 13 codes. The interviews were then coded from scratch in a two-stage process: first, coding text to all major themes in the revised codebook; then, coding these categories for subthemes (e.g., identifying types of challenges and facilitators to PCMH implementation). Team members worked in pairs on the analysis, identifying subthemes and writing summaries of the qualitative findings.

Themes analyzed from the batch of early site leader interviews include:

- reasons for participating in the FQHC APCP demonstration
- sites' experiences obtaining or planning to obtain PCMH recognition with specific discussion of PCMH component challenges
- changes in clinic structure or process expected to affect Medicare beneficiaries or other types of clients
- challenges with PCMH implementation, including specific discussion of challenges of NCQA recognition*

- facilitators of PCMH implementation, including specific discussion of facilitators of NCQA recognition*
- expected visibility of PCMH changes to patients and caregivers
- FQHC perspectives on the five intervention components offered by the CMS demonstration*
- change management strategies and ongoing changes reported by sites for PCMH implementation
- site feedback and suggestions for PCMH implementation.

Themes marked with an asterisk (*) include analysis of comparison site interviews, as well for demonstration site interviews. These comparisons allow the analyses to identify similarities and differences between demonstration and comparison FQHCs.

Three of the above themes were newly analyzed since the first annual report, including facilitators of NCQA recognition as a component of the broader discussion of facilitators of PCMH implementation, changes in clinic structure and process reported by sites as part of their PCMH implementation, and site feedback and suggestions for the demonstration. Results from the site interviews across these themes have been incorporated throughout the second annual report in order to respond closely to the current set of evaluation research questions.

Qualitative Inference and Interpretation: The qualitative sampling was designed to maximize variation of experience according to our sampling criteria (geographic region, urbanicity, and RAS score) and thus reported themes provide a range of possible alternatives rather than the most common or representative themes within the FQHC APCP demonstration or our sample of comparison FQHCs. We present all themes identified by interview respondents for a particular topic, organized by major themes with discussion of subthemes within those major categories.

PCA Leader Interviews

As described in RAND's Evaluation Design Report, we also conducted a set of baseline interviews with leaders of PCAs in each of the six states selected for the qualitative TA evaluation sample, which included three PCAs serving as cluster regional leads and three PCAs that are not. The purpose of these semistructured qualitative interviews with state PCA leaders was to learn how TA is being planned and implemented for demonstration sites. The subset of interviews with PCA cluster leads was intended to inform us about TA at both the regional and state levels. The key informants for these interviews were PCA executives and other leaders responsible for managing programs delivering TA to demonstration sites, who provided perspectives on the organization of the demonstration-related TA within the state and supplemented perspectives from the PCA focus groups with practice facilitators and coaches who interact directly with demonstration sites.

Interview topics for the PCA leader interviews included

- types of support the PCA provides to demonstration sites
- how the PCA is organizing TA to demonstration sites
- types of staff who interact directly with demonstration sites (e.g., their own staff, other PCAs, subcontractors)
- the response of demonstration sites to the TA and any issues with site participation
- the kinds of support that seem more and less helpful to sites
- main challenges that sites are having with PCMH transformation and NCQA recognition
- how the types of TA provided and experiences of demonstration sites compare with other FQHCs the PCA is supporting
- plans the PCA has for TA to demonstration sites going forward.

Interviews with lead PCAs of regional clusters included questions on coordinating TA across PCAs within their region, and perspectives on the support that the cluster lead receives from CMS and the national demonstration partners. Interviews with the other three PCAs included questions on the kinds and usefulness of support they receive from their regional cluster lead and national demonstration partners.

Analysis of PCA Leader Interviews

We used a variation of content analysis similar to the approach described for the analysis of the FQHC site interviews described above. We first developed an initial codebook based on the items in the interview protocol. A team of two coders, led by Dr. Mendel and experienced with analyses of the site leader interviews, analyzed the set of six PCA leader baseline interviews. As with the site interviews, the transcripts of the PCA leader interviews were coded in a two-stage process: first, coding text to all major themes in the codebook; then, coding these categories for subthemes if necessary, from which summaries of the qualitative findings were written.

Analyses include the following priority themes:

- differences in PCA support to demonstration versus nondemonstration sites
- barriers to providing TA to demonstration sites
- strategies for engaging sites with TA
- PCA suggestions for the demonstration moving forward.

Qualitative inference and interpretation: The qualitative sampling of state PCAs was designed to maximize variation of experience to our sampling criteria (geographic region, and leading a regional cluster of other state PCAs), and thus the reported themes provide a range of possible themes rather than the most common or representative ones within the six PCAs we interviewed. We present all themes identified by interview respondents for a particular topic, organized by major themes with discussion of subthemes within those major categories. Given the small sample of state PCA leader interviews, we do not differentiate results on the above three topics based on state PCA characteristics.

Appendix G. Site Leader Report of How Site Payments Are Used

As expected, all 20 site leader interview participants reported knowledge that their FQHC had received the per-beneficiary-per-quarter payments from Medicare as part of the demonstration.

Many interview participants did not know how the payments were used. All respondents noted that their payments were being directed to the parent organization. Most respondents were located at the parent organization level, which suggests they might know how payments were used, especially since most served in a clinical oversight and/or quality improvement position in addition to their administrative responsibilities. But more than half of the demonstration site respondents stated outright they were unaware of the details surrounding the use of the demonstration enhanced payments. Only two respondents claimed to know the exact amount of funding received. Most respondents believed these funds were used for general support of clinic operations or for changes necessary to implement the PCMH model of care. A few interviewees cited funding being directed to additional staffing. Overall, clinician respondents indicated they believed that the financial officer for their organization could provide additional information about how payments were used.

Most of the demonstration respondents described the enhanced payments as a helpful and valued support for PCMH-related changes in tightly budgeted organizations, such as FQHCs. The value of the enhanced payments to FQHCs was described as probably more significant for smaller FQHCs, but still not likely enough to sustain changes across FQHC organizations in the long term. No respondents mentioned any specific amount of payment as adequate for their purposes, even when they were specifically queried about this. However, respondents did note that the following costs associated with transformation did need to be covered: Staff for new roles (e.g., care manager, referral manager, patient educator); additional clinical staff for some sites for extended hours; maintaining IT functions and reporting for clinical process, quality improvement, and documentation; and recognition fees for recertification in the future.

One respondent, although uncertain about the amount of the payments, reported that their finance department specifically queried them about the provider transaction access numbers for the demonstration FQHC sites (within their larger, multisite FQHC organization) as a means to ensure linking the payment with that specific site. Interview participants, who were primarily clinical or operational leaders for their FQHC's PCMH efforts, typically identified the FQHC financial officer or department as the source that could best answer details on the amount and accounting of demonstration funding.

> "I could say what I think we've been using it for, but I really don't know . . . I'm not privy to it. I just know that they're letting me hire extra staff, so that's good enough for me."

Multiple interviewees noted that transformation costs occur across FQHCs within a group, while demonstration payments are applicable only to Medicare beneficiaries and only to those at demonstration FQHCs. As mentioned, respondents valued the additional funding afforded by the enhanced payments for health centers like their own that must "run tight ships."

> "I think they're helpful. When you have the size of our organization, it's not as significant as it would be in a smaller organization. But they're helpful because, again, we run very efficiently already anyway, so any little bit helps."

> "I mean, it's definitely helpful. [But our] Medicare population at the site that's participating in the project is fairly small, so it's not as if it changes our bottom line significantly."

For other demonstration FQHCs, the "additional revenue stream . . . is really not that significant," and in some cases was considered not sufficient to fully cover the changes required of the PCMH model:

> "One of the biggest challenges: We're probably going to have to hire more nonclinical staff, more in the way of medical assistants, case manager–type of roles, possibly another referrals specialist or two. And there's really not that much extra money coming in through the demonstration project to support those salaries, so we're going to have to fight a little bit to get some of that money."

None of the smaller FQHCs mentioned that payments would be more valuable for them compared to larger FQHCs, just that they valued the payments.

A demonstration FQHC site that was part of a larger set of FQHCs, including several nondemonstration sites, noted that transformation changes would have to be disseminated across all sites for the PCMH model to be sustainable within the organization. Furthermore, they suggested that if the demonstration payments are not enough to support the PCMH model in one site, they are certainly not enough to spread across all of the group's sites.

> "The project is for one of our smaller sites, but any meaningful changes that we make in the practice really have to be rolled out to all of our sites for them to be sustainable . . . otherwise, it all sort of falls apart."

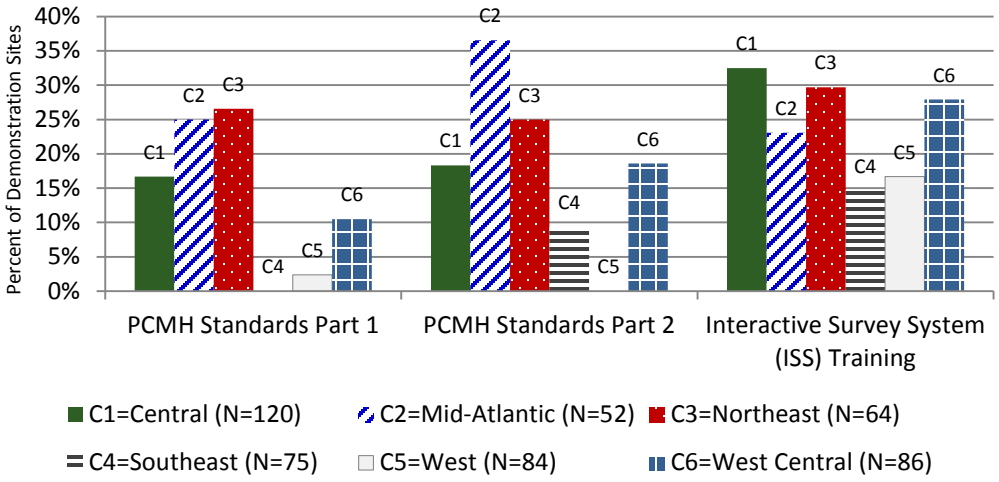
This interviewee suggested that a payment model would be needed that could support the changes throughout all grouped FQHC-sites to be sustainable.

Appendix H. NCQA Webinar Participation by Cluster

FQHC Participation in NCQA TA:
Data Source 1 of 2: RAS/NCQA Data

NCQA offers three webinars on a monthly basis—a two-part webinar on PCMH standards and a webinar on Interactive Survey System (ISS) training. NCQA developed these webinars to facilitate FQHC demonstration sites' successful achievement of Level 3 recognition. Exhibit H.1 shows cumulative participation in these webinars through June 11, 2014, by cluster. This shows that no cluster achieved a site-level participation rate of at least 40 percent for any of the three webinars. Sites in the mid-Atlantic and Northeast clusters were most likely to observe part 1 of the PCMH standards webinar, while sites in the mid-Atlantic cluster were most likely to observe part 2. Sites in the Central cluster were most likely to observe the ISS training.

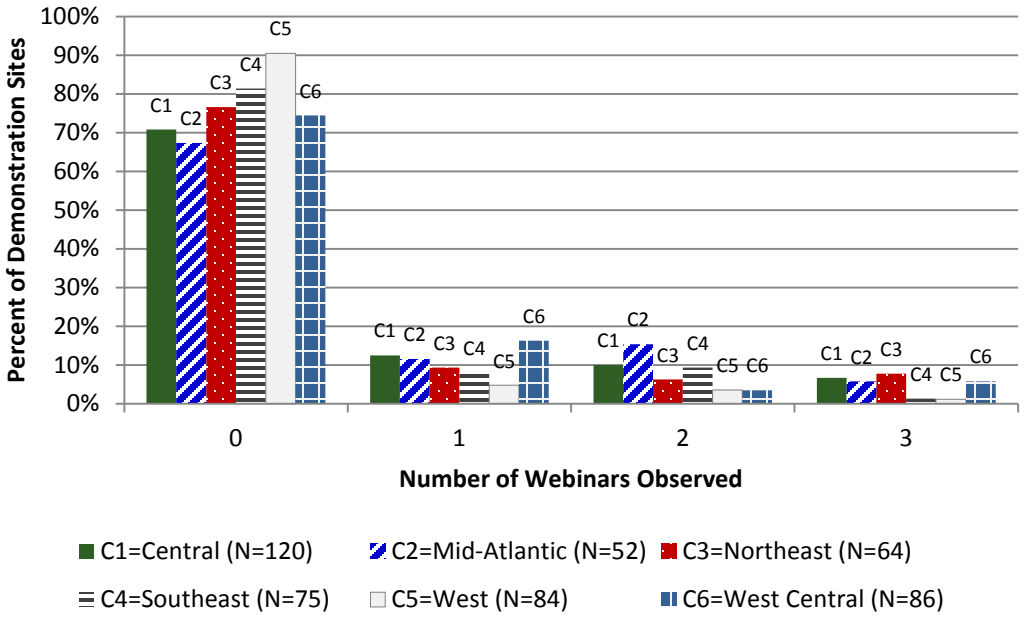
Exhibit H.1: Percent Participation in NCQA Webinars by Cluster



SOURCE: NCQA TA Participation Lists, Provided to RAND by Truven, June 11, 2014.

Exhibit H.2 displays a count of the number of NCQA webinars attended by each demonstration site. Fewer than 10 percent of sites within any cluster attended all three types of webinars, and there was limited variation across clusters in the number of webinars attended.

Exhibit H.2: Number of NCQA Webinars Observed by Cluster



SOURCE: NCQA TA Participation Lists, Provided to RAND by Truven, June 1, 2014.

Appendix I. Predictors of NCQA Webinar Participation

Among all sites, increased revenue and participating in a CMS shared-savings demonstration were both associated with a greater likelihood of participating in at least one NCQA webinar (odds ratio [OR]=1.49 for each additional million dollars of revenue and OR=1.87 for participating in a shared savings demonstration) (see Exhibit I.1). Results, however, were not always consistent across the type of TA. For example, while participating in a CMS shared-savings demonstration was associated with an increased likelihood of attending at least one NCQA webinar, it was associated with decreased odds of attending five or more AIR webinars.

Ambulatory care quality accreditation was associated with a lower likelihood of attending at least one NCQA webinar and attending five or more AIR office hours among sites that had attended any (OR=0.56 and 0.33 for NCQA webinars and AIR office hours, respectively). Similar to baseline RAS level, ambulatory care quality accreditation could be a marker for higher levels of PCMH readiness at baseline, and thus, sites that have this type of accreditation may be less dependent on TA to progress toward PCMH recognition.

Exhibit I.1: Multivariate Relationship Between Site- and Area-Level Characteristics and NCQA Webinar Participation

	Attending at least one NCQA Webinar (any type) (n = 431)	
Characteristic	**Odds Ratio**	**Standard Error**
Baseline RAS Level 0 (<35 points)	0.77	0.27
Baseline RAS Level 1 (35-59 points)	[ref]	
Baseline RAS Level 2 (60–84 points)	0.92	0.18
Baseline RAS Level 3 (85-100 points)	0.62	0.26
Receiving PCMH payments from one or more plans	0.89	0.31
Total revenue per site (in millions)	1.49*	0.20
Participation in CMS Shared Savings Demo	1.87*	0.29
Ambulatory Care Quality Accreditation	0.56*	0.24
Number of service delivery sites: 1-10 sites	[ref]	
Number of service delivery sites: 11+ sites	0.75	0.12
Rural-Urban Continuum Code: Metro	1.57	0.16
Non-metro-rural	1.97	0.22
Non-metro-urban	[ref]	
PCA Region: Central	1.71**	0.21
Mid-Atlantic	1.71	0.31
Northeast	0.94	0.27
Southeast	0.73	0.26
West	0.47**	0.25
West-Central	[ref]	

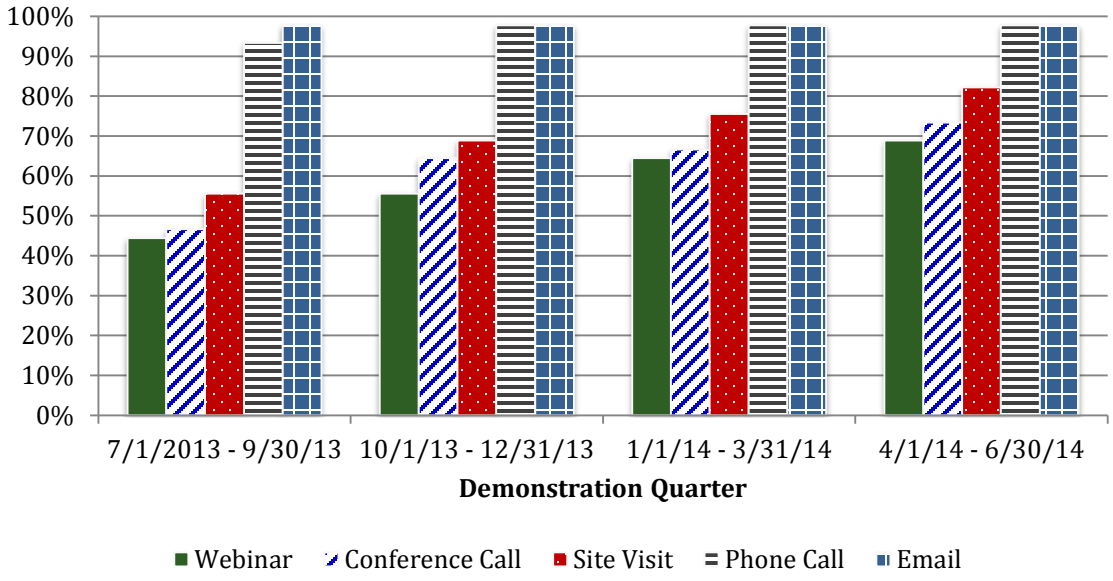
NOTE: * p<0.05; ** p<0.01; *** p<0.001

Appendix J. PCA Delivery of TA

AIR contracted with regional PCAs that in turn worked with state PCAs to structure the dissemination of TA identified by AIR. AIR and PCAs agreed to a set of reporting criteria for assessing state leadership of TA activities across the demonstration sites within their states. States reported their leadership of five different types of TA activities across four quarters beginning in July 2013.

Exhibit J.1 shows PCA-reported TA modalities led by state PCAs using reporting criteria developed jointly by AIR and PCAs, and displays the proportion of PCAs that led a TA activity for sites from July 2013 through June 2014. Across state PCAs, the most prevalent modalities for delivering TA continue to be phone calls and email. Conference calls and webinars were utilized least frequently, but increased in frequency over the year. Most recently, from April 2014 through June 2014, 32 percent of state PCAs reported not leading TA conference calls and 32 percent reported not leading TA webinars. There has been a steady increase in the use of these modalities since tracking began in July 2013. The biggest changes occurred among conference calls and webinars, where the TA modalities increased by 57 percent and 55 percent, respectively, since July 2013.

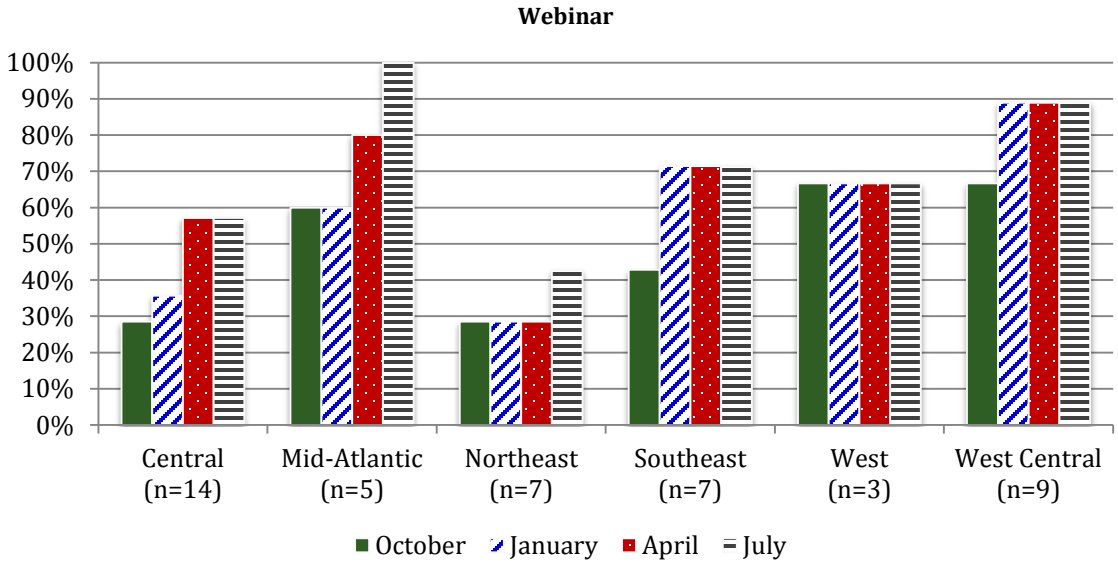
Exhibit J.1: Percentage of State PCAs Reporting Any TA Activity Over Time (n=45 State PCAs)



By late June 2014, all regions demonstrated increases in the percentage of state PCAs conducting webinars except for the West (Exhibit J.2). The mid-Atlantic region reported the

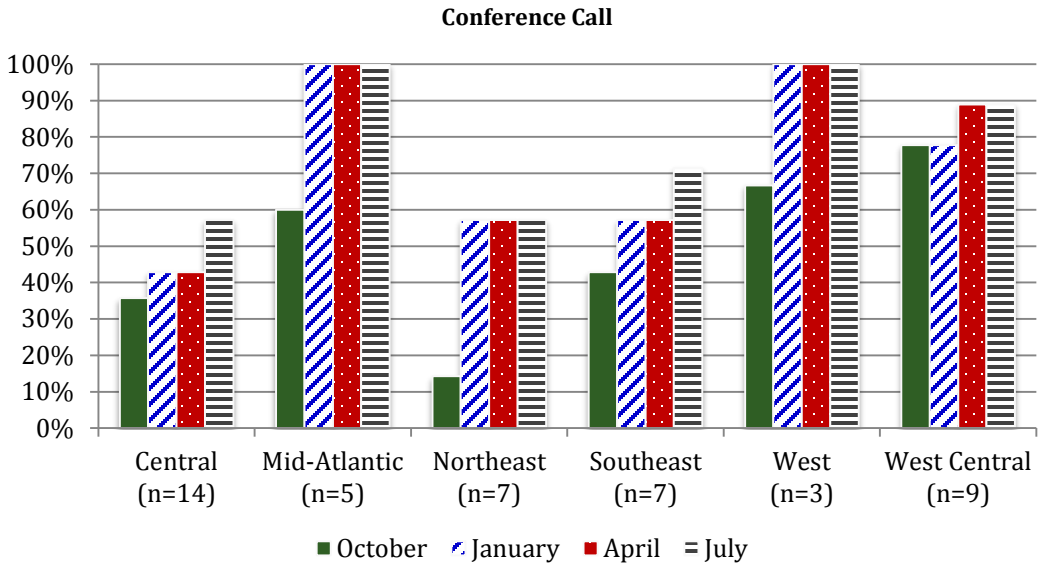
highest percent of state PCAs delivering TA by webinar leadership, while the Northeast reported the lowest.

Exhibit J.2: Percentage of State PCAs Reporting Any TA Activity by Region—Webinar



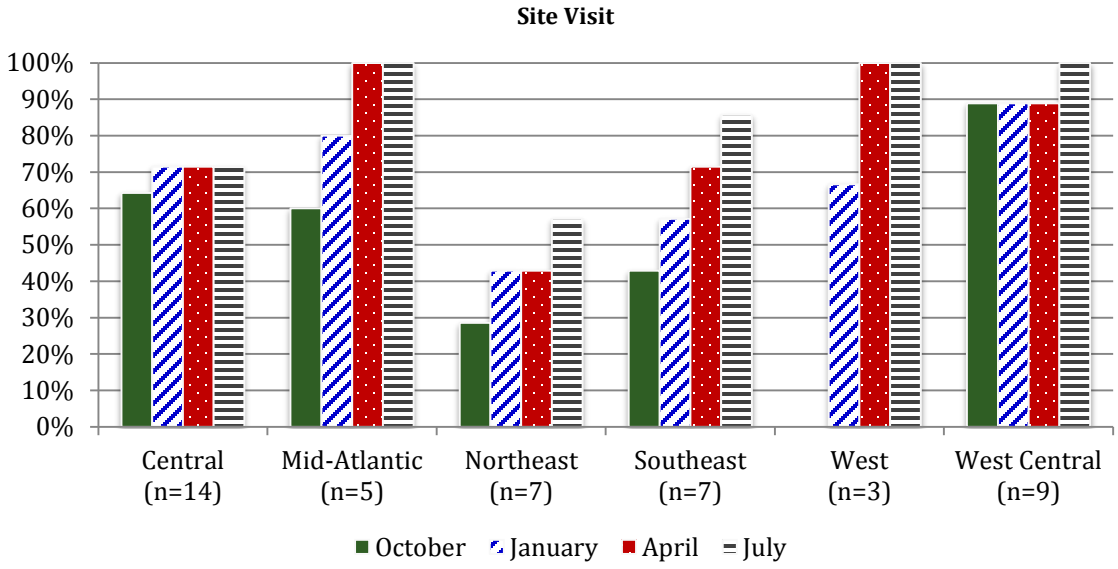
The percentage of state PCAs reporting conference call leadership increased over time, with both the mid-Atlantic and West regions achieving 100 percent activity (Exhibit J.3).

Exhibit J.3: Percentage of State PCAs Reporting Any TA Activity by Region—Conference Call



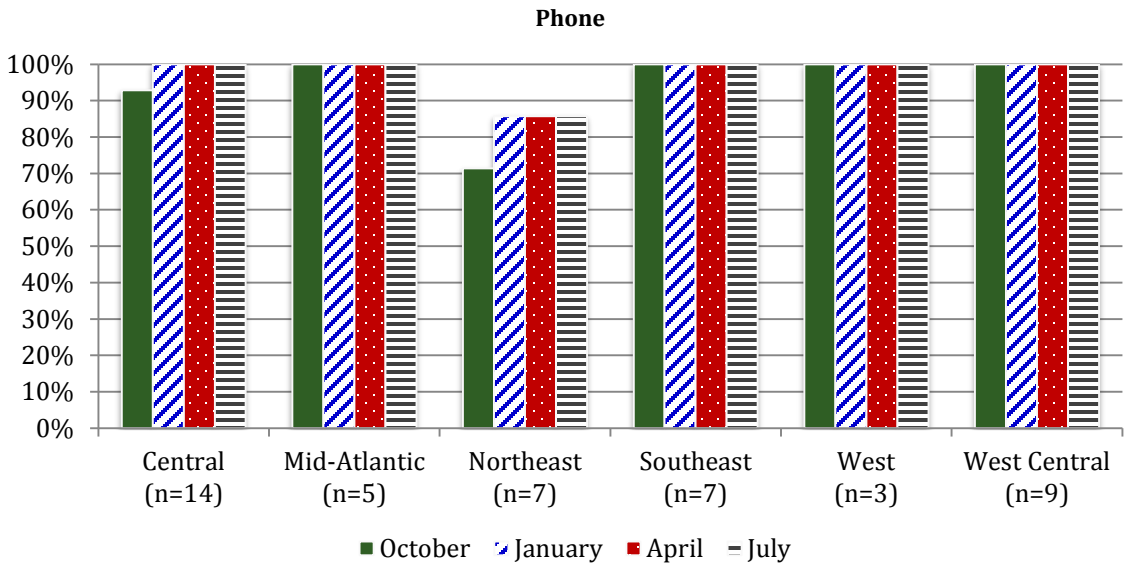
Site visits were one of three TA modalities that illustrated more variance at the aggregate level. All regions increased percentage of state PCAs engaging sites through site visits, with the mid-Atlantic, West and West-Central all achieving 100 percent of site visit activity, and the Northeast showing greatest percent change (Exhibit J.4).

Exhibit J.4: Percentage of State PCAs Reporting Any TA Activity by Region—Site Visit



Phone call activity was consistently high across regions, and all regions reached 100 percent except for the Northeast (Exhibit J.5).

Exhibit J.5: Percentage of State PCAs Reporting Any TA Activity by Region—Phone



Email activity was similarly unvaried across regions. All regions had 100 percent of state PCAs engaging sites through email over time except for the Northeast (Exhibit J.6).

Exhibit J.6: Percentage of State PCAs Reporting Any TA Activity by Region—Email