Evaluation of the Nursing Home Value-Based Purchasing Demonstration

Contract No. HHSM-500-2006-0009i/TO 7

Year 3 and Final Evaluation Report
August 26, 2013

Prepared for:
William Buczko, COR
DHHS/CMS/OA/ORDI/REG/DRTM
Centers for Medicare & Medicaid Services
7500 Security Blvd MS-C3-21-28
Baltimore, MD 21244
Phone: 410-786-6593
E-mail: William.Buczko@cms.hhs.gov

Prepared by:
L&M Policy Research, LLC
Attn: Lisa Green
1743 Connecticut Ave NW Suite 200
Washington, DC 20009
Phone: 240-476-6663
E-mail: LGreen@LMpolicyresearch.com

With Subcontractor: Harvard Medical School

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This project is the culmination of a broad and collegial effort by numerous individuals. It involved a team of researchers directed by Lisa Green, Ph.D. and managed by Julia Doherty, M.H.S.A., Rachel Dolin, Abby Hoffman, Sarah Rubin, and Sarah Rosenblum, and, from Harvard Medical School, David C. Grabowski, Ph.D., and Daryl J. Caudry, M.S..

In particular, we would like to thank our CMS Contracting Officer Representative, William J. Buczko, Ph.D. for his thoughtful guidance and support throughout the course of this project. We also would like to thank Alex Laberge, Ph.D., as well as other CMS staff, for their time and assistance with this work and Alan White, Ph.D. from Abt Associates.

Finally, the team would like to thank all of the nursing home facility administrators and staff as well as state and regional stakeholders in the demonstration states of Arizona, New York and Wisconsin who generously gave their time, perspective, and insight.

EXECUTIVE SUMMARY

The Nursing Home Value Based Purchasing (NHVBP) Demonstration is part of the Centers for Medicare & Medicaid Services' (CMS) initiative to improve the quality of care for Medicare beneficiaries in nursing homes. Launched in July 2009 and implemented by Abt Associates, Inc. (Abt), under contract with CMS, the three-year, demonstration tested the concept of value-based purchasing in nursing home settings in three states – Arizona, New York and Wisconsin. For each state, Abt ranked facilities by performance scores, which determined the annual distribution of payments. Reflecting the overall budget neutrality requirement for the demonstration, payments in each state were contingent on treatment facilities generating cost savings relative to the performance of a comparison group in each state through the reduction of avoidable hospitalizations and other costs.

CMS contracted L&M Policy Research (L&M) and its partner Harvard Medical School (Harvard) to conduct an evaluation of the demonstration that addresses the following questions:

- How does the pay-for-performance concept work within the nursing home setting?
- How has the demonstration impacted nursing home quality, cost, service delivery, resident outcomes, organizational structure, and financial status?
- How do participating homes compare to non-participating homes?

Throughout the evaluation, the L&M research team utilized a multi-method approach, integrating demonstration-generated administrative, cost, quality, and performance data with qualitative information solicited through direct discussions with nursing home staff. More specifically, this evaluation aimed to:

1. Examine organizational, patient demographic, and clinical characteristics of treatment- and comparison group nursing homes;
2. Analyze organizational, patient demographic, and clinical characteristics of nursing homes eligible for performance payments and subsequent impacts on nursing homes' quality improvement and financial status;
3. Examine the impact of the demonstration on incidence of avoidable hospitalization and quality of care levels in participating homes;
4. Assess the impact of the demonstration on nursing home management, organization, delivery of services, and financial status; and,
5. Assess the demonstration's effect on Medicare and Medicaid program expenditures and savings as well as evaluate the cost-effectiveness of the demonstration.

CMS and Abt recruited nursing homes for the NHVBP demonstration via a two-step process. First, states were asked to apply for enrollment in the demonstration. Four states – Arizona, Mississippi, New York, and Wisconsin – were selected for participation. Second, nursing homes in these four states were recruited to voluntarily enroll in the demonstration with the intent that the facilities would be randomized to treatment and comparison groups. However, New York was the only state with sufficient facility enrollment for effective randomization, with 72 facilities assigned to the treatment group. With a smaller number of facilities volunteering in Arizona (N=38) and Wisconsin (N=61), all participating facilities were assigned to the treatment

group, with a comparison group constructed using a propensity score matching process. Mississippi was excluded altogether from the demonstration due to insufficient enrollment.

For the purposes of the demonstration, each state served as a separate "laboratory" in which to test the value-based purchasing concept. Nursing home performance was assessed using measures from four domains: nurse staffing (30 percent of performance weight), quality outcomes (20 percent), survey deficiencies (20 percent), and potentially avoidable hospitalization rates (30 percent). In determining the scores for ranking performance, nursing homes accrued points based on their rank within the state. For each of the four performance domains, the lowest ranked (worst performing) nursing home received zero points. The highest ranked (best performing) nursing home each year received the maximum number of points for that measure. The exception is in the "potentially avoidable hospitalization" domain, for which the top 25 percent received the maximum number of points. For nursing homes falling between the minimum and the maximum, points were awarded proportionately based on the nursing home's rank within the distribution. Points were summed across all measures to produce an overall (composite) score for each nursing home.

A nursing home was eligible for a performance payment if it was either in the top 20 percent in overall performance, or in the top 20 percent in overall improvement *and* at least in at the 40th percentile in performance level. Those nursing homes in the top decile in any given year were potentially eligible for a higher payment than those in the second decile. Nursing homes with both high performance and improvement received the higher of the two payments. The performance payments were weighted based on the number of resident days.

The demonstration was designed to be budget-neutral with respect to Medicare. Performance payments were predicated on improvements in quality that result in a savings pool used to fund the payments. If Medicare expenditures increased by less (or decreased by more) for beneficiaries in demonstration homes relative to beneficiaries in comparison group homes, then the difference was considered Medicare savings. Importantly, in calculating potential savings, the base-year served as the comparison period for *all three* demonstration years. The distribution of any potential savings involved a shared savings approach: only the amount that exceeded the threshold of 2.3 percent of total Medicare expenditures was considered Medicare savings. Thus, savings were calculated by the difference-in-difference of pre-post spending for the treatment group versus pre-post spending for the comparison group within each state. Each year, the latter total was compared to total Medicare dollars for the demonstration group and only those dollars over 2.3 percent of total Medicare expenditures are paid out to top performing nursing homes. The size of the performance payment pool could not exceed five percent of the total Medicare expenditures, with 80 percent of the savings paid out until the five percent cap was reached and 20 percent retained by CMS. If there were no savings in a state relative to the comparison group, then no incentive payment was made to any nursing home in that state regardless of performance.

Over the entire three-year demonstration period, savings were realized in Arizona (Year 1) and Wisconsin (Years 1 and 2); no savings were generated in Arizona (Years 2 and 3), New York (Years 1-3), and Wisconsin (Year 3). Thus, only three of the nine NHVBP state-year evaluation periods resulted in payments to the top performing nursing homes. The Year 1 savings in Arizona were relatively modest, while the Year 1 and Year 2 savings in Wisconsin were more

sizable. However, in our evaluation of quality in Years 1 and 2, relatively few outcomes suggested major pre-post performance differences across the treatment and comparison groups. Based on these analyses, it appears that quality was unchanged due to the NHVBP demonstration.

In our discussions with nursing homes, officials in all three states explained that many decisions and actions, even those in quality domains targeted by the demonstration, were most likely attributable to the increasing pressures, independent of demonstration incentives, to contain costs and improve quality in response to health care reform. We heard very little to suggest that nursing homes responded to the NHVBP demonstration incentives through direct interventions.

In combining the quantitative and qualitative results, we conclude that the NHVBP demonstration did not directly lower Medicare spending and improve quality for nursing home residents. Two important questions emanate from this conclusion. First, how did Arizona (Year 1) and Wisconsin (Year 1 and 2) generate savings if nursing homes generally did not explicitly act in response to the NHVBP demonstration? And second, why did the treatment facilities appear to not respond to the payment incentives under the NHVBP demonstration?

The answer to the first question might relate to the design of the NHVBP demonstration. New York was the only state in which facilities that applied to participate were randomized across the treatment and comparison groups. Thus, the observed savings in Arizona and Wisconsin may reflect differences in facilities that comprised comparison groups selected by propensity scores in these respective states. Indeed, the difference in base-year spending for long-stay residents between the treatment and comparison facilities was much larger in Arizona and Wisconsin than in New York. Specifically, long-stay spending per day in Arizona was $15.56 (20.7%) higher in the treatment group in the base-year, while it was $6.31 (12.8%) higher in Wisconsin. By comparison, base-year spending for long-stayers in New York was $4.05 (4.8%) lower per day in the treatment group. Thus, the observed savings in Arizona and Wisconsin may simply reflect a "regression toward the mean." That is, when a variable has an extreme value on its first measurement, it will tend to be closer to the average on its second measurement.

As a sensitivity check, the treatment facilities in New York State were propensity score matched to a comparison group using the same approach utilized in constructing the Arizona and Wisconsin comparison groups. Using this alternate comparison group in New York, the treatment facilities still did not generate Medicare savings. This check is reassuring, but does not entirely rule out some behavioral effect within the treatment group in Wisconsin or Arizona. In Wisconsin and Arizona, the most engaged facilities that were interested in the demonstration and chose to participate were all assigned to the treatment group and compared against similar facilities that chose not to participate in the demonstration. In New York, the facilities that chose to participate were sorted across the treatment and comparison groups. Moreover, the NHVBP was designed such that the top performing treatment facilities received a reward payout only if the entire treatment group generated savings relative to the comparison group. A facility's likelihood of payout is based not only on their own performance but also on the performance of the other treatment facilities. Thus, the stronger the connection across the treatment group facilities, the greater the likelihood of investment in cost saving behaviors under the NHVBP. For example, if a chain enrolled all of their Wisconsin facilities in the NHVBP, this chain would have a greater incentive to invest in Medicare savings, all other things being equal. By

comparison, New York State has relatively few chains, and to the extent these chain facilities volunteered for the NHVBP, they were randomly assigned to both the treatment and the comparison groups.

Towards the second question, nursing homes may have altered their behaviors under the NHVBP demonstration for a variety of reasons. First, the demonstration had a very complex payment and reward system and nursing homes may not have understood how their efforts towards improving quality would result in a better performance score and ultimately a reward payment. Second, because CMS had a savings threshold of 2.3 percent and an 80 percent sharing rule, the payouts under the demonstration may have been too small to incentivize major changes in quality. Third, because a payout was only made if the treatment nursing homes generated savings relative to the comparison homes in that state-year period, many nursing homes may have decided not to act in direct response to the NHVBP because their likelihood of a payout depended on other nursing homes in the state also generating savings. Fourth, due to the use of administrative data to determine savings and performance, the sharing of performance reports and payouts to top performing nursing homes took nearly 18 months. This may have lowered the salience of any potential rewards to treatment facilities. Fifth, many nursing homes may have lacked the infrastructure and expertise to engage in quality improvement innovation on their own. As intended, the demonstration provided relatively little guidance and education to nursing homes as to how to improve quality. CMS had understood, perhaps incorrectly, that participating nursing homes knew what they needed to do to improve quality. The rationale for this decision was that the demonstration was designed to encourage broad innovation on the part of the participating nursing homes. Also, in thinking about the logistics of eventually launching this program nationally, it would not be realistic for CMS to provide intensive education and guidance to 16,000 nursing homes nationwide. For all of these reasons, nursing homes may have chosen to generally not respond to the quality and payment incentives under the NHVBP demonstration. Finally, rather than being incented to change practices because of the possibility of a payout, many facilities saw the demonstration as a *reinforcement of actions they were already planning to take or had already begun implementing*. Most nursing homes did not change their actions because of the demonstration; rather, some hoped to be rewarded for things that they were already doing or thought their involvement in the demonstration would just be an opportunity to learn from other nursing homes, or prepare for what is to come from CMS moving forward. Because comparison nursing homes were also undertaking similar activities to improve performance however, we did not observe a differential quality improvement in the treatment facilities.

Although the NHVBP demonstration was found to have a minimal direct effect on quality, this result may say more about the specific design features of the demonstration rather than the actual potential of nursing home pay-for-performance. If the Medicare program chooses to move forward with the pay-for-performance concept in the nursing home setting, it should consider changes to optimize the response to payment incentives to improve quality. Modifications to the design of any future NHVBP program might include: 1) simplified payment and reward rules; 2) increased payout pools; 3) relaxation/elimination of budget neutrality restrictions such that the likelihood of payout does not hinge on the efforts of other participating facilities; 4) offering more immediate payouts. 5) real time feedback on performance and quality activity results; and 6) providing increased education and guidance on best practices to providers. Towards this last

point, the program could become more prescriptive by mandating that participating providers undertake specific training or best practices in order to qualify for a reward payment.

INTRODUCTION

More than a decade after the U.S. Senate Special Committee on Aging Subcommittee on Long-Term Care (LTC) first convened in 1974 to discuss policy shortcomings in LTC, the U.S. Congress was called to focus on the quality of elderly care in American nursing homes (Institute of Medicine, 1986). A 1986 Institute of Medicine report revealed a legacy of "shockingly deficient care" across the country, and recommended that the government play a stronger role in improving nursing home quality (Turnham, 2001). Such recommendations, along with increasing public and private calls for a reevaluation of elderly care regulatory statutes, led to increased regulation and a shift away from the assumption that market forces could alone ensure an acceptable level of quality care.

Since that time—and during the last several years in particular—the nursing home industry has witnessed a number of sweeping changes aimed at improving quality in care delivery. First, in 2005, the Centers for Medicare & Medicaid Services (CMS) launched a demonstration of the Quality Indicator Survey (QIS) to test the feasibility of implementing a new survey process meant to focus on "resident-centered, outcome-oriented quality review" (CMS, 2007). Shortly thereafter, in 2006, a coalition of industry, government and consumer groups, with the support of provider organizations including CMS, launched the Advancing Excellence Campaign to encourage quality care in the nation's nursing homes. The Bush administration followed suit in 2008 when it rolled out a five-star rating system to measure and rank homes based on levels of care provided. This was intended to better aid consumers and caregivers in comparing and subsequently choosing a nursing home (HHS, 2008). In March 2010, the passage of the Affordable Care Act (ACA) dwarfed all such changes, as it included reforms that impacted all sectors of the health care system in an effort to reduce skyrocketing health care costs while improving quality. In particular, the ACA emphasized value-based purchasing and encouraged continuous quality improvements across the entire health care delivery system.

As part of this trend toward improving the quality of care in nursing homes while reducing costs, CMS contracted L&M Policy Research (L&M), LLC, and its partner, Harvard Medical School (Harvard), to conduct an evaluation of the Nursing Home Value-Based Purchasing (NHVBP) demonstration, a three-year program launched in July 2009 to improve the quality of care delivered to Medicare beneficiaries in nursing homes. More specifically, the evaluation sought to answer the following questions:

- How does the pay-for-performance concept work within the nursing home setting?
- How has the demonstration impacted nursing home quality, cost, service delivery, resident outcomes, organizational structure, and financial status?
- How do participating homes compare to non-participating homes?

This report is the third and final in a series of three annual reports on the evaluation of the NHVBP demonstration. The report explores whether a performance-based reimbursement system focusing on key quality areas may have improved the quality of nursing home care while maintaining budget neutrality, based on the data available to the evaluation team at this time. The quantitative data in this report reflects the Year 2 quality and spending results compared to the base year of the demonstration. Due to delays in obtaining Year 3 quality of care data, we are

only able to present the spending results for Year 3.[1] The qualitative findings primarily describe the perspective of participating nursing home administrators' current environment, along with general recollections of those administrators remaining at the same facility since the inception of the demonstration in July of 2009, and conversations with a number of additional administrators in Wisconsin in Year 2 from facilities that were awarded performance payments.

[1] Year 3 performance scores and other quality results are not yet available for the team's consideration at the time this report was written.

METHODOLOGY

Demonstration design

CMS and the NHVBP demonstration contractor, Abt Associates, recruited nursing homes via a two-step process. First, states were asked to apply for enrollment in the demonstration. Four states – Arizona, Mississippi, New York, and Wisconsin – were selected for participation. Second, nursing homes in these four states were recruited to voluntarily enroll in the demonstration with the intent that the facilities would be randomized to treatment and comparison groups. However, New York was the only state with sufficient facility enrollment for effective randomization, with 72 facilities assigned to the treatment group. With a smaller number of facilities volunteering in Arizona (N=38) and Wisconsin (N=61), all participating facilities were enrolled in the treatment group in each state, with a comparison group constructed using a propensity score matching process. Mississippi was excluded altogether from the demonstration due to insufficient enrollment.

In the NHVBP demonstration, each state served as a separate "laboratory" in which to test the value-based purchasing concept. Nursing home performance was assessed using measures from four domains: nurse staffing (30 percent of performance weight), quality outcomes (20 percent), survey deficiencies (20 percent), and potentially avoidable hospitalization rates (30 percent). In determining the scores for ranking performance, nursing homes accrued points based on their rank within the state. Each year, for each of the four performance domains, the lowest ranked (worst) nursing home received zero points. The highest ranked (best) nursing home received the maximum number of points for that measure. The exception was the potentially avoidable hospitalization domain, for which the top 25 percent received the maximum number of points. For nursing homes falling between the minimum and the maximum, points were awarded proportionately based on the nursing home's rank within the distribution. Points were summed across all measures to produce an overall (composite) score for each nursing home.

A nursing home was eligible for a performance payment if it was either in the top 20 percent in overall performance or in the top 20 percent in overall improvement and at least at the 40th percentile in its year's performance level. Those nursing homes in the top decile for a given performance year were eligible for a higher payment than those in the second decile. Nursing homes with both high performance and improvement received the higher of the two payments. The performance payments were weighted based on the number of resident days.

By design, the NHVBP demonstration was not prescriptive about how nursing homes should improve performance. CMS organized a series of voluntary webinars on quality improvement but facilities ultimately decided how to respond to the payment incentives inherent in the demonstration. The rationale for this decision was that the demonstration was designed to encourage broad innovation on the part of the participating nursing homes. The thinking was that—based on local economic, organizational, and policy factors—each nursing home would know best where and how to improve performance. Also, in thinking about the logistics of eventually launching a NHVBP program nationally, it would not be realistic for CMS to provide intensive education and guidance to 16,000 nursing homes nationwide.

The demonstration was designed to be budget-neutral with respect to Medicare. Performance payments were predicated on improvements in quality that resulted in a savings pool. For each state and year of the demonstration, if Medicare expenditures increased by less (or decreased by more) for beneficiaries in demonstration nursing homes relative to beneficiaries in comparison group nursing homes, the difference was considered Medicare savings. In each of the three years of the demonstration, savings were calculated relative to the base-year period (July 1, 2008 through June 30, 2009). The distribution of any potential savings involved a shared savings approach: only the amount that exceeded the threshold of 2.3 percent of total Medicare expenditures is considered Medicare savings. Thus, savings were calculated by the difference-in-difference of pre-post spending for the treatment group versus pre-post spending for the comparison group. This total was then compared to total Medicare dollars for the demonstration group and only those dollars over 2.3 percent of total Medicare expenditures were paid out to top performing nursing homes. The size of the performance payment pool could not exceed five percent of the total Medicare expenditures, with 80 percent of the savings paid out until the five percent cap was reached. Twenty percent of the savings pool was retained by CMS. If there were no savings in a state relative to the comparison group, then no incentive payment was made to any nursing home in that state, regardless of performance.

Quantitative methods

Data and study variables

This report focuses on results from Year 2 of the demonstration, which began on July 1, 2010, and ran through June 30, 2011, but also includes data from the baseline year prior to the beginning of the demonstration, (July 1, 2008 to June 30, 2009) and Year 1 of the demonstration (July 1, 2009 to June 30, 2010). For Year 3 (July 1, 2011 to June 30 2012) of the demonstration, we only examine the spending results.

First, we obtained Medicare fee-for-service eligibility and claims data from the Medicare enrollment and claims files for all individuals residing in the treatment and comparison nursing homes. Specifically, we calculated Medicare expenditures for these individuals based on their Medicare claims for skilled nursing facility (SNF) care, inpatient hospital care, outpatient hospital care, Part B (physician), and hospice care. Home health care and durable medical equipment expenditures were excluded. We only included Medicare expenditures that occurred over the course of the nursing home stay and for up to three days following the end of the stay if the individual was discharged elsewhere. To diminish the influence of cost outliers, we truncated those beneficiaries in the top one percent of Medicare expenditures in each state. Managed care enrollees and non-Medicare nursing home residents were excluded from these expenditure calculations. Finally, we calculated expenditures separately for both short-stay (post-acute) and long-stay (chronically ill) nursing home residents based on length of stay in the nursing home. Importantly, due to the shift to MDS 3.0 during Year 2 of the demonstration period, our calculation of the number of short-stay episodes changed, leading to a large increase in spending in both the treatment and comparison facilities. Because this shift happened uniformly for both groups, we do not believe it introduced any bias into the evaluation of spending in Years 2 and 3 of the demonstration.

Second, we obtained quality measures from the federally mandated Minimum Data Set (MDS) assessment instrument. MDS data are collected at the time of admission and then at least quarterly thereafter for all nursing home residents. In this study, we examine the full list of MDS-based outcomes that were used to incent performance in the NHVBP. For long-stay nursing home residents, the four measures included: the percentage of residents whose need for help with activities of daily living (ADLs) had increased, the percentage of high-risk residents with pressure ulcers, the percentage of residents with a catheter inserted and left in their bladder, and the percentage of residents who were physically restrained. The long-stay measures are the same as those reported on the CMS Nursing Home Compare tool on Medicare.gov. For the short-stay residents, the three measures were the percentage of residents with improved level of ADL functioning, the percentage with improved status on mid-loss ADL functioning, and the percentage with failure to improve bladder incontinence. As these measures were only available for the treatment facilities, they are presented in the descriptive table but not included in the regression analyses.

Third, potentially avoidable hospitalization rates were calculated using Medicare claims for both short-stay and long-stay nursing home residents. The literature suggests that a substantial portion of hospital admissions of nursing home residents can be avoided through careful management of these conditions in the nursing home. We identified those "potentially avoidable" cases as hospitalizations with any of the following diagnoses: coronary heart failure, electrolyte imbalance, respiratory disease, sepsis, urinary tract infection, and anemia (long-stay residents only). In constructing these conditions, we used both primary and secondary diagnoses. "Short-stay residents" were defined based on episodes shorter than 90 days, and we present the rate of hospitalizations per nursing home stay for this population. "Long-stay residents" were defined as individuals with a nursing home episode longer than 90 days. For this long-stay population, we constructed a measure of the number of hospitalization per 100 resident days. We included hospitalizations that occurred up to three days after the end of the nursing home stay.

Fourth, we obtained our final two performance measures from the Online Survey, Certification, and Reporting (OSCAR) system. Collected and maintained by CMS, the OSCAR data include information about whether nursing homes are in compliance with federal regulatory requirements, as every facility is required to have an initial survey to verify compliance. Thereafter, states are required to survey each facility at least every 15 months (the average is about 12 months). Nursing homes submit facility, resident, and staffing information. Deficiencies are entered into OSCAR by survey agencies when facilities are found to be out of compliance with federal regulatory standards. Each deficiency is categorized into one of 17 areas and rated by its scope and severity (on an "A" to "L" scale in order of increasing severity). In this paper, we report the total raw number of deficiencies, the number of deficiencies weighted by scope and severity, and deficiencies from complaint surveys. We also obtain staffing information from OSCAR, including registered nurses (RNs) per resident day, licensed practical nurses (LPNs) per resident day, and certified nurse aide (CNA) hours per resident day. Under the demonstration, staffing data was collected directly from payroll data submitted from treatment group nursing homes but not comparison group facilities. These data were used to calculate the score for the staffing domain. We present summary information from these payroll data for the treatment facilities as a check on the accuracy of the OSCAR staffing data. Finally, we have data on a range of potential covariates from the OSCAR, including payer mix, ownership status,

membership in a continuing care retirement community (CCRC), chain membership, hospital-based affiliation, case-mix, facility size, and location.

Statistical analysis

In evaluating nursing home performance, this study employed a "differences-in-differences" methodology, which is identified based on the pre-post difference in the introduction of the demonstration in the treatment group relative to the pre-post difference in the comparison group. The model specification is as follows:

$$Y_{it} = \beta TREAT*POST_{it} + TREAT_i + POST_t + \gamma X_{it} + \varepsilon_{it} \qquad (1)$$

in which Y is an outcome for nursing home *i* at time *t*, TREAT is an indicator for enrollment in the treatment arm of the demonstration, POST is a dummy variable for post-intervention, TREAT*POST is an interaction of the treatment and post-intervention indicators, X is a vector of covariates, and ε are the randomly distributed errors. The key parameter of interest is the interaction term between the treatment and post-intervention indicators. This model is estimated separately for each state using least squares regression. For the expenditure results, we estimate an unadjusted version of this model (excluding covariates) to mimic the approach CMS took in calculating potential savings.

Qualitative methods

The research team conducted facility and stakeholder discussions across three years, beginning in November of 2009, as the demonstration was first implemented. The first year's discussions included not only discussions with stakeholders in each of the demonstration states, but also discussions with facilities that chose not to participate. The initial facility discussions focused on why facilities chose to join the demonstration, what was involved in making that decision, beginning the data reporting, and how their facilities functioned at baseline. More information about this first year's qualitative findings was provided to CMS in the Year 1 Evaluation Report. Similar to the first year of the demonstration, the L&M team conducted two additional series of discussions in both the second and third year of the evaluation, one set with key stakeholders and another with participating nursing homes across the three states.

In this last year of the demonstration's evaluation, the L&M research team focused on particular areas of interest based on the performance data available in the spring on 2013. At that time, the second performance year data were available and reflected that Wisconsin facilities achieved a payout two years in a row. The team therefore wanted to learn more about the differences between homes that received a payout and/or demonstrated savings and those that did not. To achieve this, we added and subtracted several facilities from the discussion cohort, which allowed us to have more discussions with facilities in Wisconsin that had achieved payouts both years. It is important to note that the third and final series of discussions, while conducted once the demonstration was over, took place just as the facilities were getting their performance results (and payments) for Year 2 of the demonstration, beginning in May of 2013. Thus, the administrators were not in a position to discuss their Year 3 scores and did not know whether they were eligible for a payout in the final year of the demonstration.

Stakeholder discussions

We conducted a round of discussions with a number of the same stakeholders from Years 1 and 2 of the demonstration to refresh the valuable contextual detail about each state that these discussions provided in past years, and to refine the nursing home discussion guides. We reduced the number of total stakeholder discussions from 11 to seven for Year 3 (two from Arizona, two from New York, and three from Wisconsin), dropping from the list those stakeholders that did not previously provide key additional information. Stakeholders included representatives from for- and non-profit nursing home associations, advocacy organizations, and representatives from state Medicaid and each respective state's Quality Improvement Organization (QIO).

Facility selection

In keeping with the statement of work (SOW), the research team had been following a total of nine homes in Arizona, nine[2] in Wisconsin, and ten[3] in New York for the first two years of the evaluation period. Originally, these homes were selected to reflect the diversity of characteristics prevalent in each state, based on the baseline metrics available prior to the demonstration (e.g. number of beds, geographic location within the state, ownership type, payer mix, and quality ratings). In the second year of discussions, we included one additional home in Arizona, two in Wisconsin, and one in New York to include facilities eligible for a payout based on Year 1 results.

We conducted discussions with 30 homes during the final evaluation period, which took place after the demonstration ended, making some modifications to the list of homes from previous years. As Wisconsin was the only state to receive a payout in Year 2, we focused the majority of our resources on speaking with 20 facilities in that state. In order to do so, we reduced the number of discussions to five each in Arizona and New York. This approach sought to provide the research team additional insight into why Wisconsin homes were able to achieve savings relative to the demonstration though Arizona and New York homes were not, while still allowing us to learn about activities of the more innovative homes in the other states.

For Arizona and New York, the team removed homes from the discussion list for the final year that were in the original cohort if they did not distinguish themselves either based on their overall scores or money saved for Medicare as demonstrated by their short-stay and long-stay hospitalization rankings. Additional considerations included:

- Payer mix
- Urban vs. rural
- Ownership type (nonprofit, for-profit, or government)
- Number of certified beds
- Chain vs. non-chain (as well as whether the home is in the same chain as another home in our cohort)

[2] One of the Wisconsin homes in the original cohort dropped out after Year 1.
[3] Although the team intended to speak with nine homes in New York, one home, though listed as a single facility – and treated that way for a number of years – is legally considered to be two separate facilities with two different provider numbers.

In Arizona, few facilities in the original cohort achieved a high demonstration ranking, so we included one additional high-ranking home. For the 20 Wisconsin discussion slots, the team continued to follow homes in the original cohort – except for one home that was non-responsive in Year 2 – and added homes to be contacted that ranked high in performance and/or improvement, received a payout, or demonstrated savings through their hospitalization ranking. In addition to the cohort of ten, the team identified an additional five homes in the first cut, reserving five slots to include after contacting began. This approach would have allowed the L&M research team to make theme-based selections of the additional homes if such themes had become apparent as we delved into the discussions. Since no particular themes were identified as appropriate to guide the additional selections, we selected five more homes using the original criteria.

We selected the additional Wisconsin homes based primarily on performance score rankings. Because the original cohort included many of the lower performing homes in the state, the team added facilities ranked at the top of the state along the major scoring categories.

Table 1. Characteristics of discussants compared to all participating homes

	AZ	AZ	NY	NY	WI	WI	Overall	Overall
Characteristic	**Cohort**	**Total**	**Cohort**	**Total**	**Cohort**	**Total**	**Cohort**	**Total**
Average number of beds	131	125	187	213	95	98	116	154
Percent urban	60%	76%	80%	96%	50%	46%	57%	75%
Percent chain	60%	61%	0%	17%	35%	43%	33%	36%
Percent government	0%	0%	20%	4%	25%	14%	20%	7%
Percent for-profit	60%	76%	40%	44%	30%	50%	37%	54%
Percent Medicaid	59%	66%	64%	65%	55%	58%	58%	63%
Percent Medicare	16%	14%	13%	14%	17%	14%	16%	14%
Average Year 1 total points	52	48	51	47	42	47	47	48
Average Year 2 total points	63	48	55	47	56	50	57	48
Total number	**5**	**38**	**5**	**72**	**20**	**56**	**30**	**166**

**Cohort data is compared to all homes participating in the demonstration. Data from the comparison group nursing homes are not included here.*

Facility discussions

The team conducted a series of one-hour phone discussions from May through June 2013 using a semi-structured protocol and based on knowledge gleaned from discussions with stakeholders and findings from the previous evaluation years. Because the discussions were open-ended and

meant to solicit feedback nursing homes deemed relevant to understanding their operations as well as the impact the demonstration may have had on their facility, there was some variation in the topics and issues covered. In general, however, discussions focused on:

- Reviewing basic facility information and organizational characteristics (i.e. case mix, special units, ancillary services, staffing changes, etc.) to update responses during the previous years if already provided
- Identifying new initiatives and performance tools implemented since the previous discussions were held to include any efforts to reduce avoidable hospitalizations
- Understanding each facility's clinical staffing, relationships with physicians and hospitals
- Discussing general perceptions of the demonstration, its impact on QI and organizational activities, and any changes resulting from the demonstration.

We conducted these conversations with a combination of nursing home administrators, directors of nursing (DONs), and other staff involved with QI activities and/or data submission for the demonstration. During these discussions, we designated a note-taker, who captured responses in a note-taking template that mapped to general topics covered in the discussion guide. Where possible, we developed a series of categories that were quantifiable (i.e. yes/no answers or distinct groups) that we used in our analysis described below in the Qualitative Results section to understand the factors impacting care delivery in the nursing home setting. These categories included:

- Year 2 demonstration changes
- Administrator changes since base year
- Leader turnover since base year
- Overall reaction to the demonstration
- Eligibility for award
- Utilization of an EMR system

Note-takers recorded conversations in a transcript-like format and imported the files into Dedoose, a relational Web-based database (similar in features and functionality to NVivo and Atlas.ti) designed to support mixed-methods research. This tool allowed the team flexible data entry and analysis, including capabilities such as implementing hierarchical coding schemes and creating and modifying charts and graphics to represent quantifiable information. The tool also allows for the entire research team to simultaneously access and analyze updated versions of the qualitative discussion transcripts and the quantitative descriptors.

Limitations

Given the small number of homes per state and the qualitative basis for this portion of the evaluation, the sample size limits generalization of results to the overall sample of participating nursing homes. The research team conducted interviews for the cohort of facilities selected in the first several months of the demonstration, after the release of the Year 1 results, and following the dissemination of the results from Year 2, which occurred just a few months before the results

of Year 3 were released. As a result of these delays in the dissemination of results, the third and final set of interviews were not conducted after the demonstration was over and when Year 2 results became available.[4] By that time, administrators with whom the team spoke during each round of interviews had limited recollections about the specific timing of changes about which they were asked to report. In addition, many of the discussions conducted for this final year's report were with new administrators. As a result, the qualitative findings reported below, while providing a general sense of changes since Year 2 of the demonstration, focus more on the environment within which the facilities operated as of the writing of this report rather than the environment in place during Year 3 of the demonstration. Nonetheless, these discussions uncovered the diversity of themes and issues that likely exist for many participating facilities, and offer possible hypotheses and explanations that underlie patterns observed in the quantitative analyses.

[4] The third set of facility interviews took place between May and July of 2013.

EMPIRICAL RESULTS

This section presents quality results from Year 2 and savings results from Years 2 and 3. As this Year 3 report is also a Final Report for the evaluation, we also spend some time describing our sample and drawing broader conclusions from the entire demonstration period.

Sample characteristics

Table 2 presents baseline characteristics of the treatment and comparison nursing homes and documents the comparability of these groups, within each state. In general, there were only small differences between the treatment and comparison groups in chain membership, hospital-based status, CCRC membership, ownership type (for-profit, nonprofit, government), payer mix, case-mix, size, and location in urban areas. The treatment group in Arizona had a greater share of nonprofit facilities, hospital-based facilities, and facilities within a CCRC. Treatment facilities in New York had fewer Medicaid recipients, and treatment facilities in Wisconsin facilities were more likely to be for-profit.

Large differences existed across states in the participating homes, reflecting the different characteristics of homes within these states. For example, chain membership and for-profit ownership was highest in Arizona and lowest in New York. Wisconsin had slightly fewer Medicaid residents, more rural facilities, and lower acuity residents overall. Finally, New York had much larger facilities on average that were more likely to be located in urban areas.

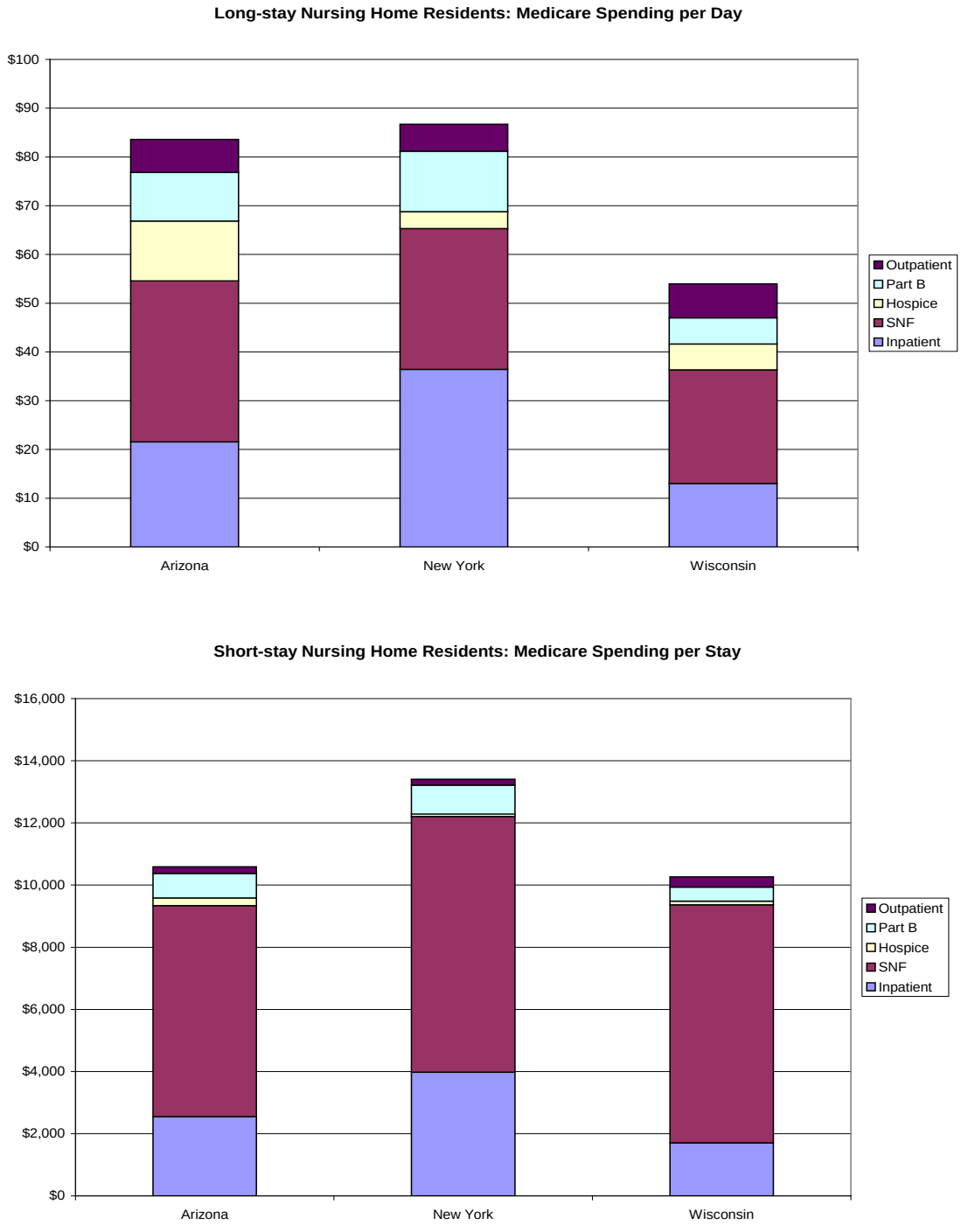
Table 2. Mean characteristics of participating nursing homes at baseline: Treatment versus comparison facilities

	AZ	AZ	NY	NY	WI	WI
Characteristic	Treatment	Comparison	Treatment	Comparison	Treatment	Comparison
Chain (%)	0.71	0.73	0.21	0.18	0.44	0.47
Hospital-based (%)	0	0.07	0.10	0.10	0.03	0.05
CCRC (%)	0.16	0.12	0.03	0.04	0.05	0.06
For-profit (%)	0.74	0.93	0.43	0.42	0.48	0.44
Nonprofit (%)	0.26	0.07	0.53	0.53	0.39	0.39
Government (%)	0	0	0.04	0.05	0.13	0.18
Medicaid (%)	0.67	0.62	0.64	0.68	0.59	0.59
Medicare (%)	0.12	0.16	0.14	0.14	0.14	0.13
Other payer (%)	0.20	0.22	0.22	0.18	0.27	0.28
Acuity score	10.55	10.31	10.61	10.60	9.53	9.62
ADL score	4.05	3.99	4.21	4.20	3.92	3.86
Total residents	97.21	95.93	198.46	197.63	89.21	83.42
Urban (%)	0.76	0.78	0.96	0.92	0.48	0.47
N	38	41	72	79	61	62

Average Medicare spending in the baseline period for long-stay residents (per day) and short-stay residents (per stay) are presented in Figure 1. Medicare spent $10,588 per short-stay episode in Arizona, $13,404 in New York, and $10,264 in Wisconsin. The bulk of the short-stay spending is driven by SNF (ranging from 61.4 percent in New York to 74.6 percent in Wisconsin) and inpatient (16.6 percent in Wisconsin to 29.7 percent in New York) services.

Medicare spends $84 per long-stay day in Arizona, $87 in New York, and $54 in Wisconsin. The major spending categories among long-stay residents are SNF (ranging from 33.3 percent in New York to 43.2 percent in Wisconsin) and inpatient (24.1 percent in Wisconsin to 42 percent in New York) services.

Figure 1. Medicare spending in baseline period by state



Changes between Base Year and Year 2

We observed some small changes across various performance dimensions between the baseline and Year 2, though no clear patterns emerged from the data analyses to support an effect of the demonstration on quality outcomes. The descriptive and differences-in-differences results are presented below.

Acuity

Two measures of resident case-mix – an acuity index and an ADL score – suggest a small increase in mean levels over the study period (see Table 3). This trend supports the broader increase in acuity that has been found within the industry as a whole over the last decade (Feng et al., 2006; Grabowski et al., 2011).

Table 3. Nursing home resident acuity across study years

Period	Acuity Measure	Observations	Mean	St. Dev.
Base Year	Acuity Index	352	10.21	1.279
Year 1	Acuity Index	352	10.27	1.336
Year 2	Acuity Index	351	10.31	1.314
Base Year	ADL Score	352	4.05	0.480
Year 1	ADL Score	352	4.10	0.482
Year 2	ADL Score	351	4.13	0.494

Quality of care

The unadjusted pre-post quality estimates for the treatment and comparison groups are presented in Table 4, Table 5, and Table 6 below. In general, relatively few outcomes suggest major pre-post quality differences across the treatment and comparison groups between the base year and year two. Outcomes that suggest better performance in the treatment group include deficiencies (Arizona, New York), long-stay hospitalizations (Arizona), long-stay pressure ulcers (Arizona), long-stay ADL worsening (Arizona, Wisconsin), long-stay restraints (New York), long-stay hospitalizations (Wisconsin) and short-stay hospitalizations (Arizona). Outcomes that suggest worse performance in the treatment group relative to the comparison group include deficiencies (Wisconsin), long-stay restraints (Wisconsin), long-stay pressure ulcers (New York) and both avoidable hospitalization measures (New York). In terms of staffing, treatment facilities in Arizona and Wisconsin performed better in RN and CNA staffing, with a relative decline in LPN staffing. In New York, the opposite was true with treatment facilities performing better in LPN staffing but worse in RN and CNA staffing. For the treatment group, the trends in the OSCAR estimates are correlated with the trends from the payroll data, supporting the accuracy of the OSCAR data.

Table 4. Arizona unadjusted differences in quality and cost outcomes: Treatment and comparison groups

Outcomes	Comparison	Comparison	Comparison	Treatment	Treatment	Treatment
	Base	Year 1	Year 2	Base	Year 1	Year 2
Deficiencies, raw count	11.63	12.88	6.90	12.42	12.97	7.32
Deficiencies, severity-adjusted score	63.46	80.34	39.00	76.26	83.29	42.63
RN hours/resident day (OSCAR)	0.34	0.42	0.47	0.32	0.39	0.47
LPN hours/resident day (OSCAR)	0.95	1.00	0.93	0.93	0.81	0.87
Nurse aide hours/resident day (OSCAR)	2.04	2.32	2.35	2.07	2.20	2.42
RN hours/resident day (Payroll)	---	---	---	0.47	0.54	0.63
LPN hours/resident day (Payroll)	---	---	---	0.86	0.84	0.82
Nurse aide hours/resident day (Payroll)	---	---	---	2.17	2.20	2.26
Long-stay ADL worsening (%)	12.73	12.71	16.31	15.14	14.28	16.7
Long-stay Pressure ulcers, high risk (%)	11.12	10.20	8.65	13.20	9.30	7.70
Long-stay Catheters (%)	6.58	5.79	6.48	5.63	6.05	5.90
Long-stay Restraints (%)	2.40	1.80	1.51	3.52	2.35	1.80

Outcomes	Comparison	Comparison	Comparison	Treatment	Treatment	Treatment
	Base	Year 1	Year 2	Base	Year 1	Year 2
Short-stay Failure to improve incontinence (%)	---	---	56.0	53.6	55.7	53.6
Short-stay ADL Improvement (%)	---	---	39.3	12.2	10.9	46.0
Short-stay Mid-Loss ADL Improvement (%)	---	---	33.2	8.7	9.6	39.2
Long-stay avoidable hospitalization rate	0.18	0.21	0.22	0.21	0.21	0.25
Short-stay avoidable hospitalization rate	0.15	0.16	0.17	0.15	0.15	0.15
Total number of homes	**41**	**41**	**41**	**38**	**38**	**38**

Table 5. New York unadjusted differences in quality and cost outcomes: Treatment and comparison groups

Outcomes	Comparison	Comparison	Comparison	Treatment	Treatment	Treatment
	Base	Year 1	Year 2	Base	Year 1	Year 2
Deficiencies, raw count	4.16	5.32	5.37	3.78	4.26	4.28
Deficiencies, severity-adjusted score	30.19	43.87	53.49	23.96	39.43	24.33
RN hours/resident day (OSCAR)	0.38	0.41	0.43	0.41	0.41	0.43
LPN hours/resident day (OSCAR)	0.81	0.81	0.80	0.81	0.83	0.84
Nurse aide hours/resident day (OSCAR)	2.19	2.22	2.26	2.31	2.32	2.34
RN hours/resident day (Payroll)	---	---	---	0.56	0.60	0.60
LPN hours/resident day (Payroll)	---	---	---	0.82	0.82	0.81
Nurse aide hours/resident day (Payroll)	---	---	---	2.24	2.28	2.24
Long-stay ADL worsening (%)	14.81	13.86	14.73	15.90	14.24	14.16
Long-stay Pressure ulcers, high risk (%)	12.88	12.49	9.56	12.44	12.22	9.8
Long-stay Catheters (%)	4.71	4.51	4.44	3.88	3.72	3.73
Long-stay Restraints (%)	2.87	2.68	2.46	2.36	2.02	1.66
Short-stay Failure to improve incontinence (%)	---	---	56.5	53.7	57.6	59.4
Short-stay ADL Improvement (%)	---	---	37.5	9.9	10.0	33.4
Short-stay Mid-Loss ADL Improvement (%)	---	---	31.6	6.4	6.5	26.8

Outcomes	Comparison	Comparison	Comparison	Treatment	Treatment	Treatment
	Base	Year 1	Year 2	Base	Year 1	Year 2
Long-stay avoidable hospitalization rate	0.19	0.19	0.22	0.19	0.19	0.23
Short-stay avoidable hospitalization rate	0.20	0.19	0.19	0.18	0.18	0.19
Total number of homes	**79**	**79**	**79**	**72**	**72**	**72**

Table 6. Wisconsin unadjusted differences in quality and cost outcomes: Treatment and comparison groups

Outcomes	Comparison	Comparison	Comparison	Treatment	Treatment	Treatment
	Base	Year 1	Year 2	Base	Year 1	Year 2
Deficiencies, raw count	7.10	6.64	6.88	5.02	6.59	6.11
Deficiencies, severity-adjusted score	67.05	47.51	54.51	38.10	47.51	57.96
RN hours/resident day (OSCAR)	0.58	0.64	0.65	0.54	0.58	0.63
LPN hours/resident day (OSCAR)	0.54	0.58	0.57	0.54	0.52	0.50
Nurse aide hours/resident day (OSCAR)	2.32	2.38	2.30	2.28	2.34	2.36
RN hours/resident day (Payroll)	---	---	---	0.66	0.79	0.82
LPN hours/resident day (Payroll)	---	---	---	0.53	0.53	0.51
Nurse aide hours/resident day (Payroll)	---	---	---	2.35	2.48	2.42
Long-stay ADL worsening (%)	14.83	13.92	14.69	16.29	13.02	15.15
Long-stay Pressure ulcers, high risk (%)	9.68	8.37	5.91	9.06	8.52	6.46
Long-stay Catheters (%)	7.09	7.51	5.61	6.10	6.03	4.72
Long-stay Restraints (%)	1.37	0.83	0.61	1.32	1.17	1.30
Short-stay Failure to improve incontinence (%)	---	---	51.6	55.5	55.7	54.1
Short-stay ADL Improvement (%)	---	---	41.9	11.1	11.3	43.9

Outcomes	Comparison	Comparison	Comparison	Treatment	Treatment	Treatment
	Base	Year 1	Year 2	Base	Year 1	Year 2
Short-stay Mid-Loss ADL Improvement (%)	---	---	35.6	9.8	9.6	38.6
Long-stay avoidable hospitalization rate	0.12	0.13	0.16	0.12	0.12	0.15
Short-stay avoidable hospitalization rate	0.14	0.12	0.13	0.12	0.13	0.11
Total number of homes	**62**	**62**	**62**	**61**	**61**	**61**

Using a differences-in-differences regression framework (see Table 7), we examined the effect of the NHVBP on a range of quality measures. In general, we have limited precision in the estimates, with only one estimate in Wisconsin suggesting a statistically meaningful ($p<.1$) result. Specifically, the severity-adjusted deficiencies score was higher in the treatment group following the demonstration. This result was also statistically meaningful in the Year 1 analyses. However, based on these results, we cannot conclude that any meaningful improvement or decline in the NHVBP Demonstration facilities relative to the comparison facilities occurred between baseline and Year 2.

Table 7. Differences-in-differences: Base and Year 2 Quality regression results

Outcomes	AZ	AZ	NY	NY	WI	WI
	Base Year	Year 2	Base Year	Year 2	Base Year	Year 2
Deficiencies, count	-0.31	(0.99)	-0.34	(0.45)	0.70	(0.63)
Deficiencies, severity-adjusted score	-5.33	(8.88)	-11.00	(7.76)	15.66*	(8.11)
RN hours/resident day (OSCAR)	0.01	(0.04)	-0.01	(0.02)	0.03	(0.04)
LPN hours/resident day (OSCAR)	-0.04	(0.06)	0.02	(0.03)	-0.04	(0.03)
Nurse aide hours/resident day (OSCAR)	-0.01	(0.13)	-0.01	(0.05)	0.05	(0.07)
Long-stay ADL worsening	-1.37	(0.88)	-0.57	(0.53)	-0.61	(0.51)
Long-stay Pressure ulcers, high risk	-1.65	(1.28)	0.16	(0.43)	0.35	(0.50)
Long-stay Catheters	0.56	(0.50)	0.43	(0.21)	0.43	(0.32)
Long-stay Restraints	-0.55	(0.38)	0.29	(0.24)	-0.02	(0.20)
Long-stay avoidable hospitalization rate	1.19	(1.30)	-0.03	(0.83)	-0.51	(0.66)
Short-stay avoidable hospitalization rate	-1.05	(0.72)	-0.28	(0.67)	-0.88	(0.55)
N	158	158	308	308	236	236

*Notes: Standard errors are presented in parentheses. * = statistically significant at 10 percent level. All regressions include the variables listed in Table 2.*

Costs

CMS estimated potential Medicare savings using an "acuity adjusted" differences-in-differences approach for each state. We follow this approach using the pre-post expenditure estimates presented in Table 8 for both Year 2 and Year 3 of the demonstration.

Year 2 Results

In Arizona, Medicare spending per day among long-stay residents increased $34.54 (or 38.67 percent) in the treatment group and $28.38 (or 38.09 percent) in the comparison group (Table 8).

Over the 218,855 long-stay resident days in the demonstration group, these findings suggest an increase in Medicare spending of $862,016. For short-stay residents, Medicare expenditures per stay decreased $889.71 (or 8.69 percent) in the treatment group and $1,301.08 (or 12.20 percent) in the comparison group. Over 5,514 short-stay episodes, the treatment facilities generated $654,450 in savings. Thus, the demonstration facilities were found to increase Medicare spending by $207,556.

In New York, Medicare expenditures per day among long-stay residents increased $33.76 (or 40.23 percent) in the treatment group and $32.41 (or 39.11 percent) in the comparison group. Over the 1,904,812 long-stay days in the demonstration group, these findings suggest an increase in Medicare spending of $1,427,332. Spending per short-stay episode increased $1,589.46 (or 12.23 percent) in the treatment group and $1,427.50 (or 10.79 percent) in the comparison group, suggesting Medicare savings of $2,988,637 over the 13,164 short-stay episodes. In total, the New York demonstration facilities generated $1,561,305 in Medicare savings.

In Wisconsin, spending per day among long-stay residents increased $19.78 (or 35.67 percent) in the treatment group and increased $24.11 (or 47.21 percent) in the comparison group, suggesting $4,166,583 in savings over the 778,031 long-stay days. Medicare spending per short-stay episode increased $703.25 (or 6.89 percent) in the treatment group and $1,172.57 (or 11.50 percent) in the comparison group, suggesting $1,834,229 in Medicare savings over 4,965 short-stay episodes. Thus, the estimated overall Year 2 Medicare savings realized by the treatment group totaled $6,000,812.

Year 3 Results

In Arizona, Medicare spending per day among long-stay residents increased $19.35 (or 21.66 percent) in the treatment group and $18.72 (or 23.94 percent) in the comparison group. The 274,330 long-stay resident days in the demonstration group resulted in Medicare savings of $557,971. For short-stay residents, Medicare expenditures per stay decreased $1,044.69 (or 10.21 percent) in the treatment group and $973.04 (or 8.8 percent) in the comparison group. Over 5,383 short-stay episodes, the treatment facilities increased Medicare spending by $777,407. Overall, the demonstration facilities were found to increase Medicare spending by $207,556.

In New York, Medicare expenditures per day among long-stay residents increased $10.73 (or 12.79 percent) in the treatment group and $11.78 (or 14.30 percent) in the comparison group. Over the 2,593,520 long-stay days in the demonstration group, these findings suggest Medicare savings of $3,299,117. Spending per short-stay episode increased $1,494.86 (or 11.5 percent) in the treatment group and $1,274.17 (or 9.74 percent) in the comparison group, suggesting a Medicare shortfall of $3,447,967 over the 15,026 short-stay episodes. In total, the New York demonstration facilities increased Medicare spending by $148,850.

In Wisconsin, spending per day among long-stay residents increased $2.71 (or 4.88 percent) in the treatment group and increased $5.51 (or 4.7 percent) in the comparison group, suggesting a Medicare shortfall of $108,887 over the 1,071,296 long-stay days. Medicare spending per short-stay episode increased $188.05 (or 1.84 percent) in the treatment group and $272.05 (or 2.67 percent) in the comparison group, suggesting $417,980 in Medicare savings over 4,934 short-

stay episodes. Thus, the estimated overall Year 3 Medicare savings realized by the Wisconsin treatment group totaled $309,093.

Table 8. Unadjusted differences in cost outcomes: Treatment and comparison groups

		Comparison	Comparison	Comparison	Comparison	Treatment	Treatment	Treatment	Treatment
State	Type	Base	Year 1	Year 2	Year 3	Base	Year 1	Year 2	Year 3
AZ	*Long-Stay Medicare Spending Per Episode*	$75.29	$81.12	$113.85	$96.95	$90.85	$93.28	$137.41	$108.68
NY	*Long-Stay Medicare Spending Per Episode*	$83.97	$85.61	$110.78	$94.11	$79.92	$83.29	$112.87	$94.65
WI	*Long-Stay Medicare Spending Per Episode*	$49.42	$52.78	$77.73	$55.92	$55.73	$53.38	$76.00	$58.15
AZ	*Short-Stay Medicare Spending Per Day*	$10,753	$10,589	$13,016	$12,037	$10,152	$9,916	$14,085	$11,281
NY	*Short-Stay Medicare Spending Per Day*	$13,122	$13,300	$13,598	$14,359	$12,740	$12,839	$13,970	$14,490
WI	*Short-Stay Medicare Spending Per Day*	$10,409	$9,848	$10,993	$10,446	$10,151	$9,831	$10,777	$10,388

Payout Eligibility

Under the rules of the NHVBP demonstration, some of the savings were shared with the highest performing facilities in the treatment group and the Medicare program retained some of the savings. In our prior year Report of the Year 1 results, we documented a relatively sizable payout to the top performing Wisconsin facilities (total payout of $4.4 million), a small payout to the top performing Arizona facilities (total payout of $33,790), and no payout to the top performing New York facilities.

In Year 2, the treatment facilities in Wisconsin generated a savings pool of roughly $6 million, of which roughly $3 million was distributed to the 17 highest performing facilities. On average, these high-performing facilities received a payment of $171,789, ranging from a low payout of $65,519 to a high payout of $361,369. In New York, the treatment facilities generated almost a half-million dollars in Medicare savings in Year 2, but because this amount was below the 2.3 percent savings threshold ($9,582,775), the top performing nursing homes in New York received no payout. Demonstration facilities in Arizona generated $207,556 in additional Medicare expenditures relative to the comparison group in Year 2. Because no savings were realized, no payouts were made to the highest performing nursing homes in Arizona.

In Year 3, the demonstration facilities in Arizona and New York had greater Medicare spending than did their respective comparison groups, while the facilities in Wisconsin generated $309,093 in savings. However, because this amount was below the 2.3 percent savings threshold of $2,618,669, the top performing facilities in Wisconsin did not receive a payout. Thus, no rewards payments were made in any of the states in Year 3 of the demonstration.

Across all three states in Year 2, we analyzed the characteristics of those top performing treatment group facilities that were eligible for a reward payout relative to the other non-qualifying treatment facilities (see Table 9). We summarized information from the base year such that any response to the NHVBP would not contaminate our findings. For example, a facility could respond to the NHVBP by admitting fewer Medicaid residents or lower acuity residents.

Table 9. Mean characteristics of top performing treatment facilities eligible for a reward payment in Year 2

	AZ	AZ	NY	NY	WI	WI
NH Characteristic	**Eligible**	**Not Eligible**	**Eligible**	**Not Eligible**	**Eligible**	**Not Eligible**
Chain (%)	60.00	60.71	13.04	18.37	33.33	47.37
Hospital-Based (%)	0	0	17.39	6.12	5.56	0
CCRC (%)	20.00	17.86	4.35	2.04	11.11	2.63
For profit (%)	60.00	82.14	34.78	48.98	22.22	62.16
Nonprofit (%)	40.00	17.86	56.52	48.98	55.56	27.03
Government (%)	0	0	8.70	2.04	22.22	10.81
Medicaid (%)	63.82	66.10	61.73	66.86	58.49	57.76
Medicare (%)	13.40	14.27	12.83	14.05	13.39	13.78
Other payer (%)	22.79	19.63	25.43	19.09	28.12	28.47
Acuity score	10.35	10.58	10.36	10.85	9.57	9.88
ADL score	3.95	4.13	4.15	4.28	4.03	4.10
Total residents	100.50	92.14	198.83	196.45	80.06	86.05
Urban (%)	70.00	78.57	95.65	95.92	44.44	47.37
Total number of homes	**10**	**28**	**23**	**49**	**18**	**38**

Notes: All variables refer to base year values. Only the top-performing facilities in Arizona and Wisconsin received reward payments. The New York treatment group did not generate Medicare savings relative to its comparison group.

Across all three states, the top-performing facilities were less likely to be for-profit owned and more likely to serve lower acuity residents. In Arizona and New York, the top performers were more likely to care for other payer (i.e., predominantly private-pay) residents. In New York and Wisconsin, independently owned (non-chain), hospital-based, and CCRC member facilities were more likely to be eligible for a reward payment.

Table 10, Table 11, and Table 12 detail base and Year 2 costs and avoidable hospitalization rates by state for short- and long-stay residents, by those facilities that qualified for a reward payment compared to those that did not. As expected of the top performing facilities, the eligible facilities generally exhibited savings and had fewer hospitalizations relative to the non-eligible facilities.

Table 10. Comparison of pre-post costs and hospitalization rates for Arizona for demonstration facilities eligible for a reward payment

Outcomes	Eligible	Eligible	Not Eligible	Not Eligible
	Base	Year 2	Base	Year 2
Short-stay cost per stay	$10,055	$11,504	$10,532	$15,007
Long-stay cost per day	$91.24	$97.80	$114.99	$151.55
Short-stay hospitalization rate	0.14	0.11	0.16	0.16
Long-stay hospitalization rate	0.18	0.18	0.23	0.27
Total number	12	10	26	28

Table 11. Comparison of pre-post costs and hospitalization rates for New York for demonstration facilities eligible for a reward payment

Outcomes	Eligible	Eligible	Not Eligible	Not Eligible
	Base	Year 2	Base	Year 2
Short-stay cost per stay	$11,502	$12,492	$12,216	$14,705
Long-stay cost per day	$69.57	$91.05	$88.22	$121.59
Short-stay hospitalization rate	0.15	0.16	0.19	0.21
Long-stay hospitalization rate	0.15	0.18	0.21	0.25
Total number	23	23	49	46

Table 12. Comparison of pre-post costs and hospitalization rates for Wisconsin for demonstration facilities eligible for a reward payment

Outcomes	Eligible	Eligible	Not Eligible	Not Eligible
	Base	Year 2	Base	Year 2
Short-stay cost per stay	$9,901	$10,029	$10,083	$11,131
Long-stay cost per day	$63.04	$71.36	$56.33	$79.67
Short-stay hospitalization rate	0.11	0.09	0.13	0.12
Long-stay hospitalization rate	0.11	0.12	0.13	0.15
Total number	18	18	43	38

Notes: All variables refer to base year values. Only the top-performing facilities in Arizona and Wisconsin received reward payments. The New York treatment group did not generate Medicare savings relative to its comparison group.

Table 13, Table 14, and Table 15 depict the differences in facilities eligible for a payout in Year 2 through an improvement score and those eligible based on overall performance by state. No consistent pattern was present as to the type of facility eligible for rewards across the three states based on the Year 2 results. It is also important to note the small number of facilities underlying the percentages in these tables.

Table 13. Comparison of descriptive characteristics of Year 2 payout eligible and not eligible facilities in Arizona

Outcomes	Not Eligible	Eligible: Improvement	Eligible: Overall Performance	Eligible: Improvement and Overall
For-profit (%)	82.14	100.00	50.00	50.00
Nonprofit (%)	17.86	0	50.00	50.00
Government (%)	0	0	0	0
Hospital-based (%)	0	0	0	0
Chain-owned (%)	60.71	50.00	75.00	50.00
Medicaid (%)	66.10	67.34	71.31	54.57
Medicare (%)	14.27	10.14	13.09	15.33
Total residents	92.14	153.50	96.75	77.75
Acuity score	10.58	10.37	10.26	10.44
ADL score	4.13	4.01	3.83	4.05
CCRC (%)	17.86	0	25.00	25.00
Urban (%)	78.57	100.00	50.00	75.00
Total number	**28**	**2**	**4**	**4**

Table 14. Comparison of descriptive characteristics of Year 2 payout eligible and not eligible facilities in New York

Outcomes	Not Eligible	Eligible: Improvement	Eligible: Overall Performance	Eligible: Improvement and Overall
For-profit (%)	48.98	37.50	33.33	33.33
Nonprofit (%)	48.98	62.50	55.56	50.00
Government (%)	2.04	0	11.11	16.67
Hospital-based (%)	6.12	25.00	0	33.33
Chain-owned (%)	18.37	12.50	11.11	16.67
Medicaid (%)	66.86	62.01	59.84	64.22
Medicare (%)	14.05	15.07	9.90	14.25
Total residents	196.45	187.88	192.44	223.00
Acuity score	10.85	10.54	10.02	10.62
ADL score	4.28	3.90	4.24	4.36
CCRC (%)	2.04	0	11.11	0
Urban (%)	95.92	100.00	88.89	100.00
Total number	**49**	**8**	**9**	**6**

Table 15. Comparison of descriptive characteristics of Year 2 payout eligible and not eligible facilities in Wisconsin

Outcomes	Not Eligible	Eligible: Improvement	Eligible: Overall Performance	Eligible: Improvement and Overall
For-profit (%)	62.16	50.00	0	20.00
Nonprofit (%)	27.03	50.00	42.86	80.00
Government (%)	10.81	0	57.14	0
Hospital-based (%)	0	16.67	0	0
Chain-owned (%)	47.37	33.33	14.29	60.00
Medicaid (%)	57.76	58.75	56.27	61.28
Medicare (%)	13.78	11.53	20.67	5.43
Total residents	86.05	74.33	76.14	92.40
Acuity score	9.88	9.69	9.63	9.33
ADL score	4.10	4.06	4.07	3.94
CCRC (%)	2.63	0	0	40.00
Urban (%)	47.37	50.00	57.14	20.00
Total number	**38**	**6**	**7**	**5**

QUALITATIVE RESULTS

Since the launch of the NHVBP demonstration in 2009, the U.S. health care industry has undergone the largest structural reform in its history. At the time of its inception, the demonstration represented a small step in the direction of reimbursement through value-based purchasing. Now, however, in the context of the passage of the Patient Protection and Affordable Care Act (PPACA), which became law on March 23, 2010, the demonstration is one of a plethora of initiatives focusing on both containing costs and improving quality and health system performance. Alongside the demonstration, such reforms, coupled with changes in reporting requirements, state economic difficulties, and other national initiatives, have created an increasingly dynamic health system that looks very different than it did in 2008.

We first present characteristics of the nursing homes with which the team held telephone discussions, followed by an overview of national changes and changes in state environments across a variety of dimensions. Finally, we present findings from discussions on the role of the demonstration on quality and outcomes, and on nursing home management and operations.

Profile of cohort

While our selected cohort represents only a small percentage of the NHVBP sample, its experiences related to reimbursement, quality initiatives, and staffing patterns largely corroborate information provided by statewide stakeholders. As stakeholders explained, many external changes have impacted internal structures and processes of the nursing homes since they enrolled in the demonstration. The following sections describe these mostly statewide changes that subsequently affected nursing home care delivery within the three demonstration states. In particular, conversations focused on health information technology, relationships between hospitals and physicians, efforts around reducing avoidable readmissions, and feedback on previous performance results from the demonstration.

Overall, homes in the discussion cohort tended to be smaller in size, located in urban areas, non-chain owned, and freestanding. They also tended to have an electronic medical record system and had neither positive nor negative feelings about the demonstration. Homes in the Arizona cohort were more likely to be urban, chain-owned, have an electronic medical record system, and have a neutral perception of the demonstration. New York homes tended to be urban and non-chain owned. Finally, the Wisconsin cohort, that represented the largest number of discussions this final year, tended to be facilities that were non-profit, non-chain owned, small, have an electronic medical record system, and have a neutral perception of the demonstration.

Table 16 below represents basic characteristics of the cohort based on Year 2 payment eligibility. In general, homes eligible for a payout were non-chain owned and either non-profit or government owned.

Table 16. Nursing home characteristics by state and eligibility for payment based on non-adjusted scores

Nursing Home Characteristic			AZ	AZ	NY	NY	WI	WI
Eligibility for Payment for Absolute Scores			Y	N	Y	N	Y	N
Bed Size	S	(30-99)	2	1	1	1	7	5
Bed Size	M	(100-159)	0	0	0	0	4	2
Bed Size	L	(160-249)	1	1	2	0	1	1
Bed Size	XL	(250+)	0	0	1	1	0	0
Urban/ Rural		Urban	3	1	2	2	5	5
Urban/Rural		Rural	0	1	1	0	7	3
Ownership		For-profit	1	2	1	1	2	4
Ownership		Non-profit	2	0	1	1	7	3
Ownership		Gov.	0	0	1	0	3	1
Within Hospital		Yes	0	0	0	0	1	0
Within Hospital		No	3	2	3	2	11	8
NH Chain		Yes	2	2	0	0	4	4
NH Chain		No	1	0	3	2	8	4
Turnover of Administrator/DON since base year		Yes	1	0	1	0	3	3
Turnover of Administrator/DON since base year		No	2	2	0	1	7	2
Turnover of Administrator/DON since base year		Not Sure	0	0	2	1	2	3
Change of Administrator since Year 2		Yes	1	1	1	0	4	3
Change of Administrator since Year 2		No	2	1	1	1	6	2
Change of Administrator since Year 2		Not Sure	0	0	1	1	2	3
EMR		Yes	2	2	1	2	8	7
EMR		No	0	0	2	0	2	1
EMR		Not Sure	1	0	0	0	2	0
Year 2 Changes Made for the Demonstration		Yes	0	0	0	1	1	0
Year 2 Changes Made for the Demonstration		No	3	2	3	1	11	8
Reaction to the Demonstration		Positive	1	0	2	1	6	2
Reaction to the Demonstration		Negative	0	0	0	0	0	1
Reaction to the Demonstration		Neutral	2	2	1	1	6	5

National Policy Changes and Initiatives

As part of the PPACA's efforts to comparison overall healthcare costs, as well as the cost to individuals, at least seven measures aimed at reducing the cost of long term care have been implemented. Three of these measures encourage increasing home and community based services and supports for residents who might otherwise be in a nursing facility. The others involve increasing access to information about nursing homes and creating a national quality improvement strategy.

As a result of section 3025 of the PPACA establishing the Hospital Readmissions Reduction Program, nursing facilities across the country have faced increased pressure from hospitals to avoid unnecessary readmissions. The program began in 2010; however, starting on Oct. 1, 2012, CMS began reducing payments to hospitals that have readmissions considered to be "preventable" under the law (CMS, 2013). Since the program was enacted, MedPAC reports that readmission rates for all conditions have declined (Lisk et al., 2013). CMS further emphasized the need to reduce readmissions by making them a part of the Quality Improvement Organization's (QIO) 2011-2014 goals (reducing readmissions within 30 days of discharge by 20 percent) (CMS, 2012).

In addition to new policies created by the PPACA, nursing homes across the country have begun using the Quality Indicator Survey (QIS), which is a revised survey system originally launched as a demonstration in 2005. CMS began training for the rollout to the first group of states, Band 1, in the summer 2009 and continued to move through each of the six bands in succession. Arizona was part of Band 1, New York was part of Band 2, and Wisconsin part of Band 4[5] (CMS, 2009). Implementation has fallen behind since, however, and several states are still transitioning to the QIS.

State Environments

Due to the different environments in each state, the effects of these national initiatives are distinct in each. In our last report, we enumerated five factors that are exemplars of the manners in which outside forces can impact nursing home care: managed care (and the extent to which it has proliferated within the state), Medicare and Medicaid reimbursements, efforts associated with reducing avoidable rehospitalizations, staffing, and Quality Improvement (QI) initiatives. After updating our literature scan and speaking with facilities and stakeholders, we have determined that these are still the most salient factors influencing nursing home care.

Arizona

Arizona's historically strong managed care environment serves to magnify the effect of each new initiative, furthering already powerful state-wide pressures to control costs. As in previous years, both stakeholders and facilities reported the need to work with both managed care organizations and Accountable Care Organizations (ACOs) in order to "stay relevant". As one Arizona stakeholder said, "If you were to [ask] in your evaluation 'has the context changed?' The answer is no, it has only increased in intensity."

[5] While Arizona now has a fully implemented QIS, New York and Wisconsin still only have it in parts of the state.

Managed Care

Arizona implemented the nation's first Medicaid managed long term services and supports (MLTSS) program in 1989, and has been on the leading edge of managed care ever since. Today, Arizona is one of two states—Texas is the other—to require Medicaid coordination with Medicare in addition to requiring contractors to offer Medicare Advantage Special Needs Plans (SNPs) (Saucier, 2012). Such coordination has allowed Arizona to lower the hospitalization rate of synchronized dual eligible beneficiaries by 31 percent. Under this model, 27 percent of dual eligible beneficiaries in Arizona deemed sick, frail, or disabled live in a SNF compared to 60 percent ten years ago (Smith, 2013). This strategy is becoming increasingly popular across the country, in part due to a new CMS demonstration, which encourages such synchronization of care for dual-eligible beneficiaries and will involve both New York and Wisconsin (CMS, 2011). As these trends continue, plan managers have increased their presence in nursing homes and, in some cases, must authorize all levels of care. State nursing home associations have made it a priority to help the plan managers and SNF administration work together. Such assistance with coordination is becoming increasingly important; one stakeholder noted that a single facility could be working with as many as 14 different managed care plans each with its own requirements.

Medicaid and Medicare reimbursement

With the new push for dual eligible beneficiaries to enroll in aligned plans, SNFs must continue to fight for reimbursement. United, one of the area's major Medicare Advantage (MA) plans, pays below the Resource Utilization Group (RUG) rates, while Aetna, another major player, pays at RUG rates. It is important for homes to be able to maintain these higher reimbursement rates in order to avoid financial stress. As noted in previous years' discussions, facilities consider Medicare beneficiaries and private pay residents highly desirable as they work towards maintaining a sound bottom line.

Avoidable Readmissions

Now that the PPACA-mandated penalties for avoidable readmissions have come into play, the pressure for Arizona homes to avoid readmissions is stronger than ever, and still surpasses that of either Wisconsin or New York. Stakeholders indicate that ACOs, particularly those that own hospitals, continue to be an increasingly important presence. As they begin to form their provider networks, ACOs direct patients to homes with which they have a partnership—either formal or informal—and SNFs must keep rehospitalizations low to have a chance of maintaining these important hospital relationships. One Arizona facility mentioned that an ACO wanted their organizations' hospitalists regularly spending time in the nursing home in order to improve quality of care. Due to this increased hospital support, this home now has full coverage from three nurse practitioners, whereas a year ago they only had one who worked part time. Many homes that we contacted in all three states said that rehospitalization is a major quality focus, although facilities in New York did not mention ACOs.

Staffing

Stakeholders and facilities mentioned minor changes in staffing, mostly related to responding to increasing acuity and decreasing rehospitalizations. Some homes said that they hired more RNs

to cover the increasingly acute patients that hospitals are discharging every year. Another mentioned that they now have a nurse exclusively to process discharges from the hospital. This is in part to ensure that the patients they do receive can be safely cared for in the SNF. Some staffing changes, however, were mentioned in relation to facilities' work with managed care organizations, resulting in the presence of more nurse practitioners at the facilities a number of times each week, keeping an eye on the residents participating in their plans. This additional perspective was generally perceived as positive as everyone was focusing more on avoiding unnecessary hospitalizations as a common goal.

Quality improvement initiatives

Many Arizona facilities mentioned the use of INTERACT II tool, particularly in working on avoidable hospitalization efforts. INTERACT II, or the Interventions to Reduce Acute Care Transfers program, provides strategies and tools to help staff notice chronically ill long-term care residents' changes in condition, with a focus on reducing avoidable hospital admissions. One stakeholder expressed that INTERACT II is still not fully integrated into their organizational culture, and therefore has not made much of an impact on facilities across the state as a whole. Another strong driver of quality initiatives in Arizona reportedly comes in the form of initiatives from individual managed care programs. Some of the plans offer bonuses to SNFs based on compliance with certain quality measures, such as ensuring that all diabetic patients receive regular foot and eye exams.

New York

Throughout this we have seen New York struggle more than the other states with Medicaid reimbursement. One stakeholder stated in this discussion round that, "The number one issue for homes is still making payroll." Another stakeholder indicated that 55 county homes have had to sell to for-profit companies because they were unable to break even. Unlike in Wisconsin and Arizona, New York stakeholders reported that most hospitals in New York are just beginning to pressure SNFs to maintain higher quality standards and focus on rehospitalizations.

Managed Care

Out of the three states included in the demonstration, New York has the least-developed managed care system. One stakeholder indicated that only the more advanced facilities are proactively partnering with managed care organizations in an effort to better position themselves in the market as managed care becomes more prominent. Another stakeholder discussed the heightened level of uncertainty surrounding managed care in New York, as it will eventually become mandatory for Medicaid recipients in nursing homes as reported in the NY Department of Health and Human Services proposal MRT #90 (NY State Department of Health, 2012). There will inevitably be major changes in New York's managed care system before then, but as of now there are few concrete details regarding this transition or the impact it will have on the nursing home industry.

Another tangible result of New York's push to increase managed care came when the state joined the CMS Dual Eligible Demonstration. New York will begin passively enrolling full dual eligible beneficiaries who require more than 120 days of facility-based long-term supports and services into a synchronized managed care plan. Beneficiaries may choose to actively enroll in a

plan beginning October of 2014, or will be passively enrolled beginning January of 2015 (NY State Department of Health, 2013).

Medicare and Medicaid reimbursement

New York has low Medicare utilization compared to Arizona and Wisconsin. New York stakeholders reported encouraging member facilities to recruit more Medicare beneficiaries in order to balance nursing homes' precarious Medicaid reimbursement. More specifically, stakeholders mentioned that it is important "to make sure [homes] have high acuity patients that Medicare pays for, so that they can survive [financially]." A recent cut in Medicaid reimbursement by four to five percentage points has exacerbated the homes' need to establish a broader payer mix to include more Medicare and private pay residents. New York homes were more likely to discuss budget cuts and financial struggles than were Arizona or Wisconsin homes.

Two of the New York stakeholders also mentioned a Medicaid global cap in the state's budget designed to limit total Medicaid spending growth. The Medicaid global cap requires greater oversight of Medicaid spending, and includes the possibility of further action if spending is projected to exceed the Medicaid cap (NY State Department of Health, n.d.).

Avoidable Readmission

Less pressure was reported by stakeholders from hospitals to avoid readmissions in New York than in the other states, and stakeholders cited several reasons that homes are falling behind on this measure. First, facilities and stakeholders alike reported strong pressure from residents' family members to hospitalize their loved ones if their health status declines, even if they could be safely stabilized in the facility. Discussions also revealed that it is general practice for clinicians to send sick patients to the hospital to avoid being responsible for a resident's deterioration, perhaps in part due to the higher incidence of lawsuits in New York than in either of the other demonstration states. As one stakeholder said, "When a nurse has to make a decision to send a patient to a hospital or not, she will send them…[since] if they have a bad outcome, their survey will crush them. If they get an IJ citation, that nurse will lose her job." Furthermore, Medicaid provides no real incentive to avoid readmissions. Medicaid will not reimburse homes for key procedures, such as administering IV antibiotics. The incentive through Medicaid is to send patients out to hospitals to complete such procedures instead of losing money by completing them in-house.

Stakeholders indicated hospitals are just starting to pressure homes to decrease readmissions because of the PPACA readmission penalties. Outside pressure from hospitals to do so was reported to be stronger in other states.

Staffing

According to stakeholders, staffing issues vary dramatically across New York State. The workforce in downstate nursing homes located near New York City is highly unionized, has more people available to fill entry-level positions, and more competition for positions requiring licensure. In the more rural northern part of the state, there is fewer staff available, and those in entry-level positions often experience transportation issues. One county-run home reported a

staffing freeze with the state requiring waivers to hire any full-time employees, leading to an increase in temporary staff in these facilities.

Quality Improvement Initiatives

New York is in the midst of developing their own statewide pay-for-performance initiative, which includes the NHVBP's long-stay hospital measures as a performance standard. This will roll out in multi-year phases. It began this year with "pay for reporting" where homes were incented for reporting certain information, regardless of their scores.

The state's QIS implementation was hampered by high survey turnover, and a lack of training dollars. Furthermore, none of the administrators mentioned the New York QIO, IPRO, as a particularly helpful resource, unlike administrators in both other participating states. One stakeholder reported that the association has been partnering with the hospital-side of the QIO to collaborate on evidence-based treatment for pressure ulcers.

There are geographic differences in the quality improvement initiatives across New York; homes downstate have had access to the Continuing Care Leadership Coalition (CCLC), an affiliate of the Greater NY Hospital Association, since 2003. The CCLC serves the New York metropolitan area but does not reach upstate facilities. It represents more than 100 non-profit and public long-term care providers throughout New York City, Long Island, and Dutchess, Orange, Rockland, and Westchester Counties (Continuing Care Leadership Coalition, 2005). In these downstate counties, the CCLC helped promote a series of projects and initiatives including the use of the INTERACT tool in facilities.

Wisconsin

Compared to both New York and Arizona, Wisconsin stakeholders and administrators alike reported a strong sense of collaboration between Wisconsin nursing homes and similarly aligned providers in the state. Consistent with Year 2 findings, stakeholders indicated that there "are not many barriers" to collaboration. Many groups from the SNF community, to the QIO, to the hospital association and the state health agencies are actively collaborating to improve quality and lower costs. This has manifested in various regional collaboratives that focus on improving transitions of care, and key stakeholder groups that "sit at the table together frequently and are very engaged."

Managed Care

Wisconsin's Medicaid managed care program, Family Care, has experienced increasing growth in enrollment throughout the demonstration. When the demonstration began in 2009, Family Care had more than 25,000 members, and over the course of the demonstration, enrollment swelled to just under 40,000 members (Wisconsin Department of Health Services, 2013). These figures include those who are enrolled in Wisconsin's Family Care Partnership Program, which serves Medicaid enrollees eligible for nursing home level of care, those dually eligible for both Medicare and Medicaid, and Program of All-Inclusive Care for the Elderly (PACE).

MetaStar's most recent report said that Wisconsin's Medicaid managed care programs "have the basic structures in place to assess and improve the quality of care" but many could "improve the

effectiveness of their quality assessment and performance improvement (QAPI) programs" (MetaStar, 2012). A continued goal for the Wisconsin Medicaid managed care programs has been to develop a more financially sustainable system.

Medicare and Medicaid reimbursement

Reductions in Medicaid reimbursement have negatively affected Wisconsin just as in the other states, but Wisconsin facilities were less likely than those in Arizona or New York to mention budget cuts as a major challenge. This may be due to the Wisconsin cohorts' larger private-pay population, which may help to balance the lower public reimbursement rates and therefore prevent homes from experiencing substantial budget cuts. For public reimbursement, according to the state association of non-profit nursing homes, Wisconsin homes are losing an average of $51.96 per day on each Medicaid resident they serve, which leads to an average annual loss of $1,125,700 per facility (LeadingAge Wisconsin, 2013). Stakeholders mentioned that trouble with Medicaid reimbursement disproportionately affects county operated homes. Wisconsin Medicaid has undergone a rate freeze. Simultaneously, residents have more acute medical problems due to increased hospital patient turnover, and, furthermore, the operating cost of each facility has increased over time.

Wisconsin stakeholders joined those of Arizona and New York in reporting the push towards Medicare and away from Medicaid in an effort to improve their financial situations. But due to federal budget cuts, even adding additional Medicare residents may not balance out homes' financial losses; in 2011, the average Wisconsin SNF experienced a Medicare rate cut of 12.6% (LeadingAge Wisconsin, 2013).

Avoidable Readmission

Stakeholders in Wisconsin mentioned that the recently announced Medicare penalties for avoidable hospital readmissions have spurred further conversations between hospitals and post-acute care providers but that some homes report only having recently focused on this issue. Still, taking into account the strong collaboration reported across interested parties, Wisconsin stakeholders as well as administrators seemed more conversant about and familiar with the steps needed to avoid readmission than New York homes. Stakeholders from Wisconsin claimed that the state has relatively low readmission rates. While they indicated there remains room for improvement, stakeholders reported that avoidable readmission rates declined since the beginning of the demonstration.

Unlike in Arizona, ACOs in Wisconsin are only in the "embryonic stage" according to one Wisconsin stakeholder. While some ACOs in Wisconsin were reported to be initiating conversations with SNFs in that state, these interactions did not appear to be as much of a driving force in increasing the focus on value based purchasing as was reported in Arizona.

Staffing

According to Wisconsin facilities, the use of NPs as more routine substitutes to physician care has become a growing trend in the state. These NPs take over the management of the physician practices when the doctors are not available. Wisconsin facilities reported utilizing more RNs in an effort to make up for the lack of available full-time physicians in the state. Similar to New

York, homes located in the rural areas of Wisconsin face greater challenges in staff recruitment and travel barriers for staff training.

Quality Improvement Initiatives

MetaStar, the Wisconsin QIO, is strongly emphasizing care transitions and many of the homes in in our last discussion cohort mentioned MetaStar meetings or summits on the subject. Wisconsin stakeholders also mentioned Advancing Excellence as a group encouraging discussions about person-centered care and rehospitalizations. Both Wisconsin stakeholders and facilities mentioned the INTERACT tool. In general, the Wisconsin homes in the cohort seem hungry for quality data and mentioned that root-cause analysis tools have been particularly beneficial.

One stakeholder mentioned a new QI tool developed by the Nursing Home Clinical Performance Measures Stakeholder group. The goal of the tool is to create a credible nursing home quality performance measurement system that can be implemented statewide to improve clinical outcomes and quality of life. LeadingAge Wisconsin has partnered with the University of Wisconsin to have a Wisconsin nursing home quality performance measurement system finalized by October 2013 (Robinson, 2012).

The QIS has not yet been fully implemented in Wisconsin due to budget constraints.

Role of Demonstration on Quality and Outcomes

In our discussions with nursing homes across all three states, and across all years of the demonstration, administrators and DONs explained that most facility changes in areas targeted by the demonstration were attributable to the increasing external pressures to contain costs and improve quality due to health reform rather than directly in response to the demonstration. A Wisconsin administrator said that the demonstration "was rewarding quality that was already being provided … reducing readmission, surveys, those were focused on initiatives that were in place anyways. Whether the payout happened or not, we would be focused on same things." One administrator in Arizona felt that the demonstration's areas of focus mirrored other measurement tools like the MDS survey, saying, "Those are things that we look at anyway. Staffing and hospitalizations are also a big focus." An administrator in New York agreed saying, "Nothing was done specifically for the demonstration; however, demonstration issues are things that are covered every month, like restraints, catheters, etc." Other homes emphasized that the demonstration reinforced internal priorities and areas of focus. A New York administrator said, "Our focus has always been quality, we didn't really do things differently because of the demo, we always had relatively low hospitalization rates. Our goals were the same as yours. We continued to try to do what we have always done." Similarly, some homes felt that the demonstration showed CMS' emphasis on the importance of quality. For a variety of reasons, administrators felt that they were already thinking about the measures emphasized in the demonstration and working on practical changes for improvement.

Some homes elected to participate in the demonstration with the hope that it would keep them abreast of approaches and perspectives on potential nursing home payment reforms moving forward. Many nursing homes noted that internal improvement and growth was a priority and they frequently participated in demonstrations and other learning opportunities. A Wisconsin administrator said, "We did it because we wanted to be the best. We always want to be better. If

[CMS is] watching these things, we're watching." Some nursing homes thought that the demonstration would give them the ability to benchmark against peers, compare best practices with other homes and gain insight from their peers. However, some were concerned that future reforms would lead to more financial stress. One administrator in Wisconsin was originally apprehensive of the demonstration and expressed the concern that by participating he would be helping CMS determine a way to pay nursing homes less.

QI activities

Most facilities reported focusing on particular areas of improvement in quality measures based on where they fall short when compared to national and state or other standards. Rather than specifically choosing to focus on the areas this demonstration highlighted, facilities reported focusing on standards developed in conjunction with corporate offices, regional associations or other organizations and modifying them internally when appropriate. Many facilities mentioned their involvement with state QIOs and localized regional collaboratives to improve quality. In part as a reflection of the statewide initiatives to improve quality and control costs in Wisconsin, most facilities that we spoke with mentioned MetaStar, the Wisconsin QIO. Many mentioned attending MetaStar events or using their tools. One Wisconsin home commented that MetaStar had been "a big driving force" behind their facility's initiatives. [6]

Administrators across all three states also discussed areas of interest mentioned in the 10th Statement of Work for QIOs, such as improving care transitions in order to reduce readmissions. Administrators reported that strengthening relationships with nearby hospitals and expanding local partnerships helped promote discussion and changes in areas such as avoidable hospitalizations. Some facilities indicated that they were able to effectively move forward with QIO goals by engaging with regional collaboratives instead of forming relationships with the QIOs themselves.

Nursing homes in all three states mentioned regional collaboratives as a resource for support in quality-related activities. One administrator in Arizona said she felt that collaborative work is the way of the future, and stayed engaged to the point that she spent time learning about pilot programs that weren't directly relevant to her facility. Some nursing homes looked to the regional collaboratives as a way to stay up to date with other local nursing homes and learn about tools or quality activities from other homes.

Facilities explained that external influences, such as local hospital activity, could also affect internal nursing home priorities. A Wisconsin administrator said, "We entered into a task force with our local hospitals to look at rehospitalization rates. We started that over a year ago. CMS had the QIO talk to us about INTERACT and MetaStar had the regional coalition, which we participated in." Some of the homes felt that as the hospitals became more proactive about

[6] Other state initiatives in Wisconsin to control costs and quality include the Wisconsin Clinical Resource Center that was mentioned in both previous evaluation years as a vital resource by many facilities. It is a joint effort between the Wisconsin Health Care Association, LeadingAge and the Wisconsin Department of Health Services, funded by civil monetary penalties collected by the state, and maintained by the Center for Health Research and Analysis at the University of Wisconsin. It offers a Web-based QI application that includes several American Medical Directors Association (AMDA) guidelines and associated training materials and is available to every nursing home in the state.

reducing readmissions, there were more opportunities to develop and improve tools to track and prevent readmissions.

An administrator in New York said that having a hospital affiliated nurse practitioner at the facility 5 days a week instead of 1 day a week gave them the clinical capacity to confirm whether there was a legitimate reason to send each patient to the hospital before doing so. Other homes redistributed resources to focus more on quality measures. One nursing home in Arizona has a specific discharge nurse, while a nursing home in Wisconsin found a physician who was also willing to review post-hospitalization data and give feedback to help decrease the potential for financial punishment based on readmission rates. That nursing home then developed a tool to track all readmissions and reasons for each readmission. The nursing home in New York that formed a value based purchasing committee in response to the demonstration developed formal processes adopted institution wide to improve collaboration and work with the nearby hospital to avoid unnecessary admissions.

QI Tools

Nursing homes continued to mention several notable tools and programs that have assisted them in their QI activities. During this year's discussions, many nursing home administrators mentioned participating in national quality initiatives such as Advancing Excellence, and using quality tools like INTERACT II. However, many of the homes clarified that they have adapted these tools to better fit their needs, either by supplementing them with their own additions or combining them with internal tools that had already been created. One nursing home in New York created a path to track anemia diagnoses based on the INTERACT diagnosis tracking tool. Another administrator in Wisconsin said, "I'm promoting it (INTERACT II) but talking to nursing staff with the initial version, they said our resident protocols or protocols for notifying physicians are already going further than what INTERACT II is telling us to do." Again, some nursing home administrators brought up their relationships with nearby hospitals when discussing QI tools. One administrator in Arizona said that a hospital near them had hosted INTERACT II training so that local administrators and DONs could all learn about it together.

Challenges associated with acuity

When discussing quality and working to improve outcomes, many administrators also discussed the challenges associated with the trend of admitting patients with higher levels of acuity into nursing homes. This trend has forced some nursing homes to adapt by making changes in staffing and tailoring training to accommodate higher acuity patients. One nursing home in New York said that they have encountered a lot of difficulty due to increasing acuity, particularly in the evening when there is less support than during day shifts. The administrator at this home felt that patients were often sent out of the hospital too early, and they often had no choice but to send them back. One stakeholder indicated that homes are often at the mercy of hospitals, and must admit any patients sent to them in order to maintain a good relationship. Many homes mentioned that having mid-level providers or physicians on staff helped them to reduce their readmissions. Some indicated that they needed to change staffing ratios due to the ever-increasing acuity. An administrator in WI explained that the facility was very proud of their readmission rate given the high acuity of their patients, saying,

> *"We take very ill people and we take patients that others can't handle. We feel that the biggest thing that keeps our rehospitalizations low is that we are fortunate enough to have two MDs on staff that round Monday-Friday. Whenever one is off and the other is back on, they do a complete check-in with each other. The RNs all have a unit manager; they know their patients well. The RNs and MDs have a unique relationship, so if the RN calls the MD in the middle of the night, they know each other and have a good idea of whether or not the patient needs to go to the hospital or can be treated in the home. Our CNAs also have a lot of longevity, and they know the patients well."*

Besides having a well-trained staff, this facility felt that good communication skills and familiarity with the patients helped keep very acute patients from returning to the hospital. One administrator in Arizona said that 26 of the 30 total nurses employed by the facility were RNs. A state contact with Arizona said that, though LPNs were still the most common nurses in LTC, some nursing home administrators might skew towards RNs due to the higher acuity in the managed care environment. Finally, a different nursing home in Arizona said that they wanted to be able to accept patients with a higher level of acuity but they would need to hire a higher-trained staff and increase the number of ancillary services offered. Because this would require substantial investment, they want to make sure that their local market can support those departments before making this decision.

Similar to the situation mentioned in New York, a nursing home in Arizona felt that some families still pressured nursing homes to send relatives to the hospital. They said, "Often when someone starts to fail, the family jumps on that immediately. We work really hard with the residents that we know."

Role of Demonstration on Management and Operations

As with QI efforts, when asked to describe the reasons for changes in their operations and care delivery practices since the inception of the demonstration, almost all administrators in this year's discussions indicated that the demonstration, at best, played only a small factor in altering operations. Many administrators suggested that while the goals of the demonstration were generally consistent with their overall goals in terms of improving internal operations and staffing, the pressures to do so would be there regardless of the demonstration's existence.

One notable exception was the same New York home mentioned in previous years' reports. They were not eligible for a payment, but nonetheless indicated that the demonstration definitively resulted in that facility developing a new quality initiative focusing on areas highlighted by the demonstration. They formed a value-based purchasing committee in response to the demonstration that is now a standing committee reviewing all hospitalizations to ensure any hospitalizations that can be appropriately avoided are. They indicated that their participation in the demonstration led them to focus on rehospitalization issues long before they would have otherwise, and "put them well ahead" of similar facilities in New York now that this measure is becoming a major focus across the country.

Turnover in management

While many nursing homes in the cohort had not experienced recent turnover in management positions, administrators hired after the demonstration started frequently felt that they did not have a clear understanding of the demonstration. Many were unaware that their facility was even involved in the NHVBP demonstration until they received a payout, or a phone call from the evaluation team asking to schedule time for a discussion. One administrator who joined the facility after the start of the demonstration said that he first heard about the demonstration a few months after he started "looking at information and data, you could see in financial data that there was a larger sum of revenue in one month…that is when I learned about demonstration project and that that was a reward." When a nursing home was not directly involved in the data-submission process, such as when the headquarters of a chain submitted data for multiple facilities, administrators were less likely to be aware of the demonstration. In many cases, when demonstration results were sent to chain corporate offices, administrators were unaware of their facility's performance results or scores until they spoke with the evaluation team.

Labor and staffing

As mentioned previously, many staffing changes described by administrators were attributed to issues around increasing acuity and other increased documentation requirements. The demonstration may have had a marginal impact on some facilities' staffing, since facilities were tracking additional staffing information for the demonstration and possibly paid more attention to their staffing levels and turnover than they might have otherwise. One facility in Wisconsin said that many of their seasoned nurses had retired, and while the newer nurses often had less experience, they expected that more acute patients could be cared for in the nursing home setting. Another home in Wisconsin said that they had changed hiring practices based on turnover rates, saying, "We aren't hiring brand new CNAs anymore, and I think that's why people like this are basically set up to fail."

Other facilities mentioned factors outside of their control, such as geography and the local marketplace, factored into recruiting and retention. Some of the more rural facilities said that there was often a smaller pool of applicants, while facilities in more saturated markets found that competition between facilities could make hiring more difficult. Another facility in Wisconsin said that they had tried to hire fewer part-time people to help build levels of trust and continuity between patients and staff and also among staff members. Other facilities used RNs or NPs to supplement physicians and offer more supervision to patients.

Facilities also mentioned additional access to physicians and mid-levels from hospitals or managed care plans.

Capital Investments

Similarly, many changes in capital appeared to be attributable to environmental pressures at the national and state level. Several homes reported changes designed to appeal to the Medicare population and private pay population, such as creating more private rooms to attract younger populations, specifically Baby Boomers, coming to skilled nursing facilities. Other homes discussed a growing interest in rehabilitation units and services.

Many facilities not only invested in some form of Electronic Medical Record (EMR), but also discussed their ongoing efforts to make sure that the EMR was being utilized to its full potential. A few homes were using their EMRs with limited functionality, but many of those were either in the process of fully implementing it or planning a full implementation. One facility in Wisconsin was integrated into their local hospital's EMR, EPIC, and was able to use this tool to give input on patients' readiness for discharge. Most facilities had implemented an EMR system to better track patients and improve their management reporting capacity.

Administrators continued to mention the increasing availability of ancillary and other specialty services in nursing homes. Many nursing homes mentioned offering ancillary services on site and for longer hours or on weekends as a vehicle to decrease avoidable hospitalizations and safely take in patients with higher acuity levels. One nursing home in Wisconsin said that they were able to pull patients from a larger geographic area because they can administer intravenous (IV) therapy. A nursing home in Arizona runs stat labs and other ancillary services at the facility, and credits this with avoiding hospitalizations. As more acute patients are discharged from hospitals, many nursing homes have continued to add more highly trained staff and more ancillary services to not only compete for such patients, but also to feel secure that they can provide necessary and appropriate care.

Data Reporting

As noted in the previous report, administrators did not mention any significant burden associated with the reporting required specific to the NHVBP demonstration once they had their systems set up in the first year.

Nursing home administrators reported mixed feelings, however, about the MDS survey and their respective performance scores under the demonstration. Some felt that the rankings were unclear and were not entirely sure how their NHVBP performance scores were derived. Others felt that the MDS survey itself was punitive and did not look at facilities fairly. One facility in Wisconsin, which had many patients with brain injuries or mental illness, said,

> *"If you are taking residents with few behaviors, your antipsychotic numbers will be low. For us, we take so many mental health issues, that our antipsychotic numbers are higher. These medications allow these patients to participate in therapy and live more normal lives… the MDS doesn't take into account mental issues."*

They also mentioned that the MDS survey didn't account for baseline measurements regarding brain injury and only looked for decline.

Several administrators mentioned that they would like to see more benchmarking capabilities or be able to learn best practices from facilities that scored in the top ten percent within the demonstration each year.

Appropriate documentation as a way to increase quality scores emerged as a theme across states. One Wisconsin administrator credited his facility's improved scores in the NHVBP demonstration to their significant investment in teaching staff how to use their EMR to its full potential and accurately coding MDS, rather than any real change in quality of care. Another facility in Wisconsin hired a nurse consultant to focus on documentation, saying, "That could be

a factor that may have influenced our scores this year. She was looking at short stay residents and change of condition documentation…I think that picking that up earlier has allowed us to really find at-risk residents before they go to the hospitals." One facility in Arizona expressed frustration that their scores were not as good as some of their peers, and attributed their lower performance to their inability to document accurately. Finally, some administrators saw the MDS survey and scores as a good starting point for conversations on areas that needed to be improved.

DISCUSSION

Although some nursing home administrators mentioned the burden of data collection in the first year of the demonstration, most felt that it became much simpler and less time consuming once there was a system in place. Though the demonstration quality reports were long out of date by the time they were distributed, administrators often appreciated having access to these data. Most indicated they had hope for far more frequent and current information to be provided to them throughout the demonstration to allow for benchmarking against other like participating facilities. For some homes, the CMS reports were the only source of these benchmarking-type quality metrics, so they found them useful despite the fact that they were retrospective. One facility even mentioned using data from the demonstration showing their high ranking within the state as they approached ACOs to present a case for inclusion on their preferred provider list. In general, homes were interested in not only benchmarking, but many would have liked to receive more information about best practices and other suggestions for how to continuously improve around the demonstration performance metrics.

Several homes, regardless of their eligibility for an award, mentioned a feeling of unpredictability about receiving a payout. One administrator hypothesized that his good scores were a more likely the result of the luck his facility had in the performance year, since they had not interacted with many families pushing for hospital admission that year and had not had many patients' conditions worsen severely. Others felt that they were excelling in the quality measures and were therefore confused when they did not rank high enough to be eligible for a payment. The lack of regular and timely data made it difficult for homes to determine how well they were performing in the demonstration. Because they received outdated information once a year that was sometimes as much as 18 months old, they were not only wholly unable to predict their outcomes, but also unable to target specific areas for demonstration improvement during each respective demonstration year.

One stakeholder summed up the feelings of many involved in the demonstration saying, "This was an absolute missed opportunity." This stakeholder felt that their association had a lot to offer in terms of leadership, but was only enlisted to help recruit homes. Associations in each of the states also mentioned they were not consistently invited to listen in on the quarterly NHVBP demonstration calls. They were hoping for more information about the progress of the participating facilities to use as a learning opportunity for their constituencies throughout the life of the demonstration. Nursing home administrators had hoped for more communication about their overall progress as well as best practices. In general, administrators and stakeholders felt that the demonstration was a good idea, but lacked the necessary communication and leadership to really impact quality measures.

Finally, some administrators expressed confusion at the algorithms behind the scores, as well as frustration at the inconsistencies between their ranking in the demonstration as compared to other rankings, such as their 5 star ranking on Nursing Home Compare. One administrator in Arizona said, "We've participated and sent in everything they asked us to but to be honest I was never 100% sure what they were looking for and then to hear we didn't qualify just makes me think what didn't we have what [they] were looking for?" This uncertainty surrounding the demonstration was not helped by staffing turnover at top levels within many of the facilities, as

many administrators who oversaw the demonstration for Year 3 were new and not the administrators (or their corporate offices) who had decided to participate in the demonstration.

Rather than being incented to change practices because of the possibility of a payout, many facilities saw the demonstration as a reinforcement of actions they were already planning to take or had already begun implementing. Most nursing homes did not change their actions because of the demonstration; rather, some hoped to be rewarded for things that they were already doing or thought their involvement in the demonstration would just be an opportunity to learn from other homes, or prepare for what is to come from CMS moving forward. Thus, the qualitative analyses indicated very little direct effort on the part of demonstration facilities towards improving quality and lowering Medicare expenditures in direct response to the demonstration.

The quantitative analyses also suggested very little response to the NHVBP payment incentives. In our analyses of quality in Years 1 and 2, we found very little impact of the demonstration on nursing home performance in the treatment facilities when compared to comparison facilities. Over the entire three-year demonstration period, savings were observed in Arizona (Year 1) and Wisconsin (Years 1 and 2). Put alternatively, no savings were found in Arizona (Years 2 and 3), New York (Years 1-3), and Wisconsin (Year 3). Thus, only three of the nine NHVBP state-year evaluation periods resulted in payments to the top performing nursing homes. The Year 1 savings in Arizona were relatively modest (total payout pool was roughly $27,000), while the savings in Wisconsin were more sizable (the total payout pool was roughly $3.5 million in Year 1 and $3 million in Year 2).

Lessons learned

In synthesizing the quantitative and qualitative results, we conclude that the NHVBP demonstration did not directly lower Medicare spending and improve quality for nursing home residents. Two important questions emanate from this conclusion. First, how did Arizona (Year 1) and Wisconsin (Year 1 and 2) generate savings if nursing homes generally did not explicitly act in response to the NHVBP demonstration? And second, why did the treatment facilities appear to not respond to the payment incentives under the NHVBP demonstration?

The answer to the first question might relate to the design of the NHVBP demonstration. New York was the only state in which facilities that applied to participate were randomized across the treatment and comparison groups. Thus, the observed savings in Arizona and Wisconsin may reflect differences in facilities that comprised comparison groups selected by propensity scores in these. Indeed, the difference in base-year spending for long-stay residents between the treatment and comparison facilities was much larger in Arizona and Wisconsin than in New York. Specifically, long-stay spending per day in Arizona was $15.56 (20.7%) higher in the treatment group in the base-year, while it was $6.31 (12.8%) higher in Wisconsin. By comparison, base-year spending for long-stayers in New York was $4.05 (4.8%) lower per day in the treatment group. Thus, the observed savings in Arizona and Wisconsin may simply reflect a "regression toward the mean." That is, when a variable has an extreme value on its first measurement, it will tend to be closer to the average on its second measurement.

As a sensitivity check, the treatment facilities in New York State were propensity score matched to a comparison group using the same approach utilized in constructing the Arizona and

Wisconsin comparison groups. Using this alternate comparison group in New York, the treatment facilities still did not generate Medicare savings. This check is reassuring but it does not rule out some behavioral effect within the treatment group in Wisconsin or Arizona. In Wisconsin and Arizona, the most engaged facilities that were interested in the demonstration were all assigned to the treatment group and compared against similar facilities that chose not to participate in the demonstration. In New York, these engaged facilities were sorted across the treatment and comparison group. Moreover, the NHVBP was designed such that the top performing treatment facilities received a reward payout only if the entire treatment group generated savings relative to the comparison group. A facility's likelihood of payout is based not only on their own performance but also on the performance of the other treatment facilities. Thus, the stronger the connection across the treatment group facilities, the greater the likelihood of investment in cost saving behaviors under the NHVBP. For example, if a chain enrolled all of their Wisconsin facilities in the NHVBP, this chain would have a greater incentive to invest in Medicare savings, ceteris paribus. By contrast, New York State has relatively few chains, and to the extent these chain facilities volunteered for the NHVBP, they were randomly assigned to both the treatment and the comparison groups.

Towards the second question, nursing homes may have altered behaviors under the NHVBP demonstration for a variety of reasons including:

- *Complex Payment and Reward System:* Incentive-based payment systems work well when providers understand how effort links to performance and ultimately to a reward payment. The NHVBP had a very complex payment and reward system based on a number of measures and relative and absolute performance. Nursing homes may not have understood how their efforts towards improving quality would result in a better performance score and ultimately a reward payment.

- *Small Reward Payments:* The size of the potential reward payment will inevitably influence the response by providers. CMS had a savings threshold of 2.3% and an 80% sharing rule to ensure that any savings were not due to simple statistical noise but rather reflected true savings on the part of the participating nursing homes. Thus, the payouts under the demonstration may have been too small to incentivize major changes in quality. Indeed, over $8 million in "savings" was retained by CMS under the shared savings rules.

- *Role of External Factors:* A well-designed incentive system minimizes the role of external factors outside the facility's control in determining the likelihood of a reward payment. Under the demonstration, a payout was only made if the treatment nursing homes generated savings relative to the comparison facilities in that state-year period. Thus, many nursing homes may have decided not to act in direct response to the NHVBP because their likelihood of a payout depended on other nursing homes in the state also generating savings.

- *Payment and Information Lag:* Real-time payouts ensure that facilities can recoup investment in quality improvement relatively quickly. However, due to the use of administrative data to determine savings and performance, payouts to top performing nursing homes took up to 18 months. This may have lowered the salience of any potential rewards to treatment facilities. This lag in payment and the corresponding lag in

information that would otherwise allow facilities to get ongoing feedback as to their performance during the demonstration did not encourage the facilities to benchmark against themselves or their peers on the demonstration measures throughout the demonstration.

- *Lack of Education and Guidance:* Many researchers have argued that incentive payments may not work well in the context of complicated tasks such as improving nursing home performance. The idea is that poor performance relates to both misaligned payment but also a lack of on-the-ground knowledge on how to improve performance. The demonstration was designed to address the misaligned incentives, but nursing homes may still have lacked the infrastructure and expertise to improve performance. As intended, the demonstration provided relatively little guidance and education to nursing homes as to how to improve quality. The rationale for this decision was that the demonstration was designed to encourage broad innovation on the part of the participating nursing homes. Also, in thinking about the logistics of eventually launching this program nationally, it would not be realistic for CMS to provide intensive education and guidance to 16,000 nursing homes nationwide.

- *Facilities were already doing quality interventions:* Rather than being incented to change practices because of the possibility of a payout, many facilities saw the demonstration as a *reinforcement of actions they were already planning to take or had already begun implementing*. Most nursing homes did not change their actions because of the demonstration; rather, some hoped to be rewarded for things that they were already doing or thought their involvement in the demonstration would just be an opportunity to learn from other nursing homes, or prepare for what is to come from CMS moving forward. Because comparison nursing homes were also undertaking similar activities to improve performance however, we did not observe a differential quality improvement in the treatment facilities.

- All of these explanations may have contributed to the limited quality improvement and savings found under the demonstration. This result may say more about the specific design features of the NHVBP demonstration rather than the potential of nursing home pay-for-performance more generally. If the Medicare program chooses to move forward with the pay-for-performance concept in the nursing home setting, it should consider changes to optimize the response to payment incentives to improve quality. Modifications to the design of any future NHVBP program might include: 1) simplified payment and reward rules; 2) increased payout pools; 3) relaxation/elimination of budget neutrality restrictions such that the likelihood of payout does not hinge on the efforts of other participating facilities; 4) offering more immediate payouts. 5) real time feedback on performance and quality activity results; and 6) providing increased education and guidance on best practices to providers. Towards this last point, the program could become more prescriptive by mandating that participating providers undertake specific training or best practices in order to qualify for a reward payment.

WORKS CITED

CMS. (2007). *Evaluation of the Quality Indicator Survey (QIS)* (Contract No. 500-00-0032, TO#7). Retrieved from https://www.cms.gov/CertificationandComplianc/Downloads/QISExecSummary.pdf.

CMS (2009, August 7). *Priority Order of Quality Indicator Survey (QIS) National Implementation in States (Ref: S&C-09-50).* Retrieved from http://www.cms.gov/SurveyCertificationGenInfo/downloads/SCLetter09_50.pdf

CMS (2011, April 14) *15 States Win Contracts to Develop New Ways to Coordinate Care for People with Medicare and Medicaid*. Retrieved from http://www.cms.gov/apps/media/fact_sheets.asp

CMS (2012, May). *Quality Improvement Organization: Current Work*. Retrieved from http://www.cms.gov/Medicare/Quality-Initiatives-Patient-AssessmentInstruments/QualityImprovementOrgs/Current.html

CMS (2013, August 2). *Readmission Reduction Program.* Retrieved from http://cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html/

Continuing Care Leadership Coalition. (2005). *The evolving need for long term care services in New York State* [Powerpoint slides]. Retrieved July 8, 2013, from http://www.leadersincare.org/images/CCLCChartBook.pdf

Feng, Zhanlian; Grabowski, David C; Intrator, Orna; Mor, Vincent (2006, August). The Effect of State Medicaid Case-Mix Payment on Nursing Home Resident Acuity. *Health Services Research*. doi: 110.1111/j.1475-6773.2006.00545.x

Grabowski, D. C., & Town, R. J. (2011). Does Information Matter? Competition, Quality, and the Impact of Nursing Home Report Cards. *Health Services Research*, 46(6pt1), 1698-1719. doi:10.1111/j.1475-6773.2011.01298.x

Institute of Medicine, Committee on Nursing Home Regulation. (1986). *Improving the Quality of Care in Nursing Homes*. Washington, D.C.: National Academy Press. Retrieved from http://www.nap.edu/openbook.php?record_id=646.

LeadingAge Wisconsin. (2013, March). Wisconsin nursing facilities: Medicaid and family care deficits for payment year 2011-12. Retrieved from http://www.leadingagewi.org/sites/default/files/iplosses.pdf

Lisk, David et. al (2013, March). *Refining the Hospital Readmissions Reduction Program.* MedPac. Retrieved from: http://www.medpac.gov/transcripts/readmissions%20Public.pdf

MetaStar. (2012, November). External quality review report Wisconsin Medicaid managed long-

term care family care, family care partnership and program of all-inclusive care for the elderly state fiscal year 2011-2012. Retrieved from http://www.dhs.wisconsin.gov/ltcare/StateFedReqs/eqro11-12.pdf

New York State Department of Health (2012, September). *Report to the Governor and the Legislature.* Retrieved from http://www.health.ny.gov/health_care/medicaid/redesign/docs/mltc_legislative_report.pdf

New York State Department of Health (2013). *Addendum to NYSDOH's May 24, 2012 Demonstration Proposal to Integrate Care for Dual eligible Individuals and Request for Public Comment.* Retrieved from http://www.cms.gov/Medicare-Medicaid-Coordination/Medicare-and-Medicaid-Coordination/Medicare-Medicaid-Coordination-Office/Downloads/NYAddendum_forpubliccomment.pdf

New York State Department of Health. (n.d.). *Monthly and regional global cap updates*. Retrieved from http://www.health.ny.gov/health_care/medicaid/regulations/global_cap/

Robinson, J. (2012, December). A Wisconsin nursing home clinical performance measurement system -- Options and issues. Retrieved from http://www.wahsa.org/sites/default/files/chsrareport.pdf

Saucier, Paul; Kasten, Jessica; Burwell, Brian; Gold, Lisa (2012, July). *The Growth of Managed Long-Term Services and Supports (MLTSS) Programs: A 2012 Update*. Centers for Medicare and Medicaid Services. Retrieved from http://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Delivery-Systems/Downloads/MLTSSP_White_paper_combined.pdf

Smith, Susan (2013, March). *Arizona's Managed Care Succeeds in Aiding Vulnerable, Low-Income Patients.* Wolters Kluwer Law & Business. Retrieved from http://health.wolterskluwerlb.com/2013/03/arizonas-managed-care-succeeds-in-aiding-vulnerable-low-income-patients/

Turnham, H. (2001). *Federal Nursing Home Reform Act from the Omnibus Budget Reconciliation Act of 1987.* National Long Term Care Ombudsman Resource Center.

U.S. Department of Health and Human Services. (2008). *Trends in Nursing Home Deficiencies and Complaints* (OEI-02-08-00140). Retrieved from http://oig.hhs.gov/oei/reports/oei-02-08-00140.pdf.

Wisconsin Department of Health Services (2013, June). *Family Care Enrollment Data.* Retrieved from http://www.dhs.wisconsin.gov/ltcare/Generalinfo/enrollmentdata.htm

APPENDIX

Discussion guides - Qualitative data collection
Nursing Home Value Based Demonstration Project

Demonstration Year 3 Key Informant Telephone Discussions

Key Stakeholders (associations and regulators in 3 states)

Spring/Summer 2013

Procedures for obtaining informed consent during telephone discussions and recording

Prior to beginning the call, tell them that we normally record discussions to help with note taking. Confirm and document that they agree to have the call tape-recorded. State that we will not associate their name or the name of their facility when we report our findings.

Overall Goal

The overall goal of the project is to assess the impact of the CMS-sponsored Nursing Home Value Based Purchasing Demonstration on nursing homes' quality of care. This is the continuation of a three-year assessment using a combination of methods to include discussions with nursing home staff, and analysis of data from the demonstration and from administrative files such as OSCAR, to better understand how the concept of "pay for performance" works in the nursing home setting.

Introduction

Welcome

Thank you for agreeing to speak with us today.

My name is {NAME, title} from L&M Policy Research, and I am calling you on behalf of the Centers for Medicare and Medicaid Services (CMS). L&M is a health policy research firm located in Washington, D.C., that CMS has contracted with to conduct an ongoing assessment of the Nursing Home Value-Based Purchasing Demonstration.

Background—explain purpose of the call

The purpose of our discussion today is to get your perspective as we update our environmental scan of the nursing home industry in [state] and to learn of any changes that may impact the facilities participating in the demonstration.

Everything you tell us will be confidential in that we will not connect your name or the name of your organization with anything you say.

Ground rules

- Our discussion will be between 45 minutes and an hour.
- Because we are recording please try to speak in a voice as loud as the one I am using now so that we can make sure the tape is picking up our conversation.
- Questions?

Introduction

Just to get acquainted [or reacquainted], please tell me a bit about yourself and your organization's activities with nursing homes and/or related to the demonstration in [state].

- Tell us what your role is, if any, in the demonstration [or if it has changed in any way]. PROBE: "Formal" versus "informal" involvement.
- Has the role your organization played during the implementation of the demonstration changed since we last spoke?

Updating the Environmental Scan

- Are there any new specific regulatory changes unique to [state] of which we need to be aware? If so, why?
    - PROBE: State Medicaid changes? Other changes in reimbursement or state surveys? (Implementation of QIS or other surveyor changes?) Changes in bed hold policies?
- Are there any additional specific challenges that you consider unique to your state? If so, what are they? [List those they mentioned last year if they do not mention them again, and ask if they are still relevant, such as managed care and ACO activity, etc..]
- Are there any specific challenges that might be unique to certain counties or areas of the state in comparison to other areas?
- Please describe any P4P initiatives in the long term care arena that have been implemented since 2010 in your state, and whether these initiatives have changed nursing home perspectives on P4P...

[If they indicate any regulatory or other environmental changes of importance, ask for them to refer us to the relevant documents, provide links, or email us copies.]

Perceptions of the Demonstration

- Has your organization had questions or concerns about the demonstration over the last three years, and if so, did you get your questions answered?

- Do you think the demonstration impacted overall nursing home care for participating facilities? If so, how and why?

- What structural and operational changes are you aware of that nursing homes have made in response to the demonstration in the past three years?

- What kinds of questions and feedback did you receive from nursing homes in your state about the demonstration, if any? Did the nature of those questions or feedback change over the course of the demonstration?

- What did you generally find to be the participating nursing home staff's perceptions about the demonstration? Were those perceptions different than those at non-participating homes? Have any of those perceptions changed throughout the course of the demonstration?

IF NOT ALREADY COVERED:

- Did any of the participating facilities talk to you about their performance scores and rankings in the past year?

- What do you believe are the most relevant challenges nursing homes are facing at this point in time related to the goals of the demonstration?

    - Are these challenges the same or different from those in previous years?

    - Are there any other contextual factors we should understand while interpreting and understanding our discussions with facilities this year?

CLOSING

Before we end, I'd like to give you chance to share any additional thoughts or comments about what is going on in your state related to the nursing home industry and/or with the demonstration. Is there anything else you would like to add that you didn't have a chance to say during our discussion today?

Thank you very much for participating in this discussion today. We appreciate you taking the time to speak with us.

Discussion guide - Qualitative data collection
Nursing Home Value Based Demonstration Project

Demonstration Year 3 Key Informant Telephone Discussions

Treatment Nursing Facility Group

Summer 2013

Procedures for obtaining informed consent during telephone discussions and recording

Prior to beginning the call, tell them that we normally record discussions to help with note taking. Confirm and document that they agree to have the call tape-recorded. State that we will not associate their name or the name of their facility when we report our findings.

Overall Goal

The overall goal of this project is to assess the impact of the CMS-sponsored Nursing Home Value-Based Purchasing Demonstration on nursing homes' quality of care. This is the third set of discussions we are conducting as part of a three-year assessment. [Indicate whom you spoke with at their facility before for the year 2 discussions if this is not the first call.] We are using a combination of methods including discussions with key stakeholders, nursing home staff, and analysis of data from the demonstration and from administrative files such as OSCAR, to better understand how the concept of "pay for performance" works in the nursing home setting.

INTRODUCTION

Welcome

Thank you for agreeing to speak with us today.

My name is {NAME, title} from L&M Policy Research, and I am calling you on behalf of the Centers for Medicare & Medicaid Services (CMS). L&M is a health policy research firm located in Washington, D.C., that CMS has contracted with to conduct an ongoing assessment of the Nursing Home Value-based Purchasing Demonstration.

Background—explain purpose of the discussion

- Our discussion is part of CMS' Nursing Home Value Based Purchasing initiative.
- The purpose of our discussion today is to focus on your facility's experiences with the demonstration.
- Everything you tell us will be confidential in that we will not connect your name or the name of your facility with anything you say.

Ground rules

- Our discussion will last about an hour.
- Because we are recording please try to speak in a voice as loud as the one I am using now so that we can make sure the tape is picking up our conversation.
- Questions?

Introduction

Just to get acquainted, please tell us a bit about yourself and your facility [and if this person participated in the discussion last year, also ask about any significant changes in your facility in the past year].

- Tell me your job title and your role at [name of nursing home facility.]
- Tell me a little bit about your facility [Mention information garnered from previous discussions or research and ask them to confirm its accuracy].
    - Probe key characteristics such as patient mix (acuity); ownership; staffing.
    - Determine whether the facility includes units for dementia, ventilator or dialysis dependent, or other special patient categories.
    - Determine whether the home has ancillary services, such as a pharmacy, X-ray, or medical staff readily available onsite 24/7 or on weekends.
- Tell us a bit about the environment within which your facility operates and the extent to which it has changed since last year.
    - Probe on survey environment, payments by state, other quality initiatives in which the facility participates or is contemplating participation.

NURSING HOME FACILITY: BACKGROUND

Describe what we know about their facility from previous years' discussions, if they were previously part of the nursing home discussion cohort, AND based on the information collected as part of the Year 3 facility selection process.

Tell us more about your facility. PROBE: Has your mission, philosophy or your goals as a skilled nursing facility changed in the past several years? Do you offer any services and programs not available three years ago?

- Tell us about your management team and how your facility is organized. How has it changed in the past two years?
- In general, how do you approach staffing (clinical staffing and more broadly)?

DECISION-MAKING: PARTICIPATION IN THE DEMONSTRATION

- [If not already determined from previous discussions] Tell us about your perceptions of the demonstration.
    - PROBE: Do you recall why your facility decided to join the demonstration? Did it meet your expectations?
    - PROBE: Was there information you would like to have received which might have been more helpful in the implementation of the demonstration?
- How do you think the concept of "pay for performance" will (or currently) work(s) in your facility in general? Has there been a consensus among decision makers in your facility on this?
    - PROBE: (If there is no consensus) What have others been saying?
- Have you seen your performance rankings and report for Year 2? If so, please tell us what you thought of it, and how your facility fared.
    - PROBE: Was the report clear and understandable?

CLINICAL CARE DELIVERY MODEL (QI, PERFORMANCE MGT.)

- Tell us about your current model of clinical care delivery. Did the demonstration influence your clinical care delivery at all? What changes have taken place that you can attribute to the demonstration?
- Did your nursing home already use any performance-oriented tools or incentives prior to implementing this demonstration? Did this change due to the demonstration? If so, how?
    - Does your facility participate in the Advancing Excellence program or any quality or process improvement initiatives other than those you've already mentioned?
- Tell us about your current quality improvement plan. To what extent has the demonstration impacted your quality improvement activities?
- What, if any, changes did your facility make for the demonstration? PROBE: Overall staffing changes? Operational (changes in policies and procedures)? Structural changes?
- Did you target certain areas for improvements or special focus given the outcome measures in place for the demonstration? If so, which?
- Do you have any programs in place to address avoidable hospitalizations? If so, please tell us how these programs came to be. To what extent have these programs been implemented as a result of the demonstration?

- What were your facility's goals with respect to your quality performance score? Have they changed since the inception of the demonstration?
    - PROBE: On which, if any, of the performance score categories did you focus: staffing performance measures; hospitalizations; MDS scores; or survey deficiencies?
    - Tell us how you have worked to meet these goals and your experience doing so, including successes or challenges you have faced.
- Which aspects of performance do you find you can best influence? Why? And what areas in terms of performance has your facility struggled with the most? Please explain.
- How much staff time and resources did you dedicate to demonstration reporting on a day-to-day basis after the initial reporting systems were set up?
- What do you think the overall impact of the demonstration has been on your facility/system with respect to cost savings? PROBE: Profitability?
- *IF CHANGES IN OVERALL CLINICAL CARE DELIVERY, ORGANIZATIONAL STRUCTURE DISCUSSSED:* How do you think the changes your facility made played out? To what extent were they in response to the demonstration? PROBE: Short-term? Long-term? Are these changes sustainable? Why or why not?

DEMONSTRATION IMPLEMENTATION

- Tell us overall about your facility's experience with the implementation of the demonstration. PROBE: Burdensome? Occurred as expected?
- What questions, if any, have you and your staff had since beginning your participation in the demonstration? If you had any questions, where did you go to get them answered? Who provided this information and how helpful was it?
- Did you participate in the CMS quarterly calls throughout the demonstration? Were your expectations met? Were there areas where you would have liked additional assistance/clarification?

CLOSING

Before we end, I'd like to give you chance to share any additional thoughts or comments about the demonstration.

- Is there anything else you would like to add about the demonstration or the topics we have covered already that you didn't have a chance to say during our discussion today?

Thank you very much for participating in this discussion today. We appreciate you taking the time to speak to us.

L&M POLICY RESEARCH