© Published 2017. This article is a U.S. Government work and is in the public domain in the USA
DOI: 10.1111/1475-6773.12688
RESEARCH ARTICLE

A Longitudinal Assessment of the Effect of Resident-Centered Care on Quality in Veterans Health Administration Community Living Centers

Jennifer L. Sullivan, Michael Shwartz, Kelly Stolzmann, Melissa K. Afable, and James F. Burgess

Objective. To examine whether changes in resident-centered care (RCC) over time were associated with changes in quality.

Data Sources/Study Setting. Data sources were the Minimum Dataset quality indicators (which consist of measures of both prevalence and incidence of adverse events) and the Artifacts of Culture Change Tool (which measures RCC; FYs 2009–2012) from 130 Veterans Health Administration community living centers.

Study Design. A retrospective longitudinal study.

Data Collection/Extraction Methods. Data were from VA secondary data sources.

Principal Findings. The overall relationship between RCC and quality was not statistically significant ($p = .22$), although there was a weakly significant negative relationship (i.e., increased RCC was associated with poorer quality) in the seven quarters after implementation of an automated version of the Artifacts Tool ($p = .08$). In facility-specific analyses, there were 15 facilities with a weakly significant ($p < .10$) positive relationship between RCC and quality and 21 with a weakly significant negative relationship. Adjusted cost per patient day was over 50 percent higher in the 21 facilities with a negative relationship than in the 15 facilities with a positive relationship ($p < .05$).

Conclusions. The Artifacts score is a formal performance metric in the VA, and thus, facilities were explicitly incentivized to increase RCC. Using qualitative methods to identify characteristics that distinguished those facilities able to increase both RCC and quality from those that suffered declines in quality as RCC was improved is an important follow-up to this study.

Key Words. Nursing homes, resident-centered care, quality, Veterans Health Administration

Following the passage of the Omnibus Budget Reconciliation Act of 1987, skilled nursing facilities (SNFs) became focused on a medical model of care, with an emphasis on standardization, safety, and clinical quality (White-Chu

et al. 2009). Since the mid-1990s, several models of culture change that emphasize resident-centered care (RCC) have been implemented. These models involve shifting the care paradigm to promote resident quality of life and resident self-direction. In particular, RCC includes individualized care, transformed physical environments, and changed staff roles (Ronch 2003; Rahman and Schnelle 2008). Grabowski et al. (2014) theorized that implementation of RCC would positively affect both resident quality of life and, in turn, clinical quality because they are complementary. For example, RCC would result in more meaningful staff–resident relationships, which could in turn lead to earlier identification of quality problems such as pressure ulcers.

Studies examining the effect of RCC on quality of care have reported a modest positive association (Stone et al. 2002; Kane et al. 2007; Mueller 2008; Grabowski et al. 2014; Shier et al. 2014). However, previous studies are limited in that they examine different types of RCC practices, and the studies typically are conducted in small settings, hindering generalizability (Port, Sloane, and Zimmerman 2005; Grant and McMahon 2008; Rahman and Schnelle 2008; Afendulis et al. 2016). In the first large-scale longitudinal study of RCC and quality of care in 215 exemplar SNFs that had already adopted RCC, a positive relationship between RCC and quality was not found, although there were no statistically significant declines in quality over time (Grabowski et al. 2014). Nevertheless, there remains a gap in the empirical literature regarding how the degree of RCC adoption is related to quality of care in a large group of facilities varying on the extent of RCC implementation.

The Veterans Health Administration (VA) is a large national integrated health care network that has adopted RCC throughout its system and renamed its SNFs to community living centers (CLCs) as part of this initiative. The implementation of RCC in the CLC setting was developed and launched in 2005 and was included in the VA 2006 strategic plan. The standardized dissemination of RCC precepts in VA CLCs, and their measurement by a standard tool (i.e., Artifacts of Culture Change Tool), provided an opportunity to

Address correspondence to Jennifer L. Sullivan, Ph.D., Center for Healthcare Organization and Implementation Research, VA Boston Healthcare System, 150 S. Huntington Ave (152M), Boston, MA 02130 and Boston University School of Public Health, Boston, MA; e-mail: Jennifer.sullivan@va.gov. Michael Schwartz, Ph.D., is with the Center for Healthcare Organization and Implementation Research, VA Boston Healthcare System, Boston, MA, and Boston University Questrom School of Management, Boston, MA. Kelly Stolzmann, M.S., and Melissa K. Afable, M.P.H., are with the Center for Healthcare Organization and Implementation Research, VA Boston Healthcare System, Boston, MA. James F. Burgess, Ph.D., is with the Center for Healthcare Organization and Implementation Research, VA Boston Healthcare System, Boston, MA, and Boston University School of Public Health, Boston, MA.

examine the relationship between the extent of RCC implementation and quality of care as measured by the prevalence and incidence of adverse events. A cross-sectional study using these data found higher quality in facilities with greater levels of culture change (Sullivan et al. 2013). A more recent longitudinal study found no overall change over time in rates of pressure ulcers after the implementation of RCC in VA (Hartmann et al. 2016). The goal of our study is to extend the cross-sectional analysis of Sullivan et al. (2013) by examining the relationship between extent of RCC implementation and a broad-based measure of clinical quality longitudinally within facilities. We hypothesize that higher RCC implementation scores over time would be related to better quality of care (i.e., composite Minimum Dataset quality indicator scores).

METHODS

We conducted a retrospective longitudinal study from Fiscal Year (FY) 2009 to 2012 using facility-level data from 130 VA CLCs. This study was approved by VA's Central Institutional Review Board.

Measures

RCC Implementation Score. The Veterans Health Administration uses the Artifacts of Culture Change Tool to measure the extent of RCC implementation on a quarterly basis. Beginning with the last assessments in FY 2010, VA began to use an automated system for capturing Artifacts Tools data to be consistent with the Center for Medicare & Medicaid Services (CMS) scoring approach (see Sullivan et al. 2013). However, leadership at VA only selected artifacts which were applicable to the VA. Thus, many environmental artifacts were removed from the Tool. For the purposes of this study, we rescored the Artifacts Tool responses from the end of FY2010–FY2012 to be consistent with the scoring approach used by VA in the prior years.

The Artifacts Tool has six domain scores: care practices (45 points), workplace processes (55 points), environment (180 points), leadership (25 points), family and community (30 points), and outcomes such as longevity and turnover (20 points). The domain scores are summed together to create a total Artifacts score (up to 355 points). Data from the fourth quarter of FY2009 and the second and fourth quarter of FY2010 were unavailable. For unclear reasons, the VA did not process data from those time periods, most

likely due to internal changes in staffing. We imputed the total Artifacts Total Score for the three missing quarters as the average of the quarter that preceded and followed the missing quarter.

MDS Composite Measure. Our dependent variable was quality of care as measured at the facility level by a composite measure created from the 28 Minimum Dataset (MDS) 2.0 quality indicators (QIs). As the MDS QIs measure unfavorable events happening to residents (e.g., development of pressure ulcers), lower rates indicate higher quality. Similar to Sullivan et al. (2013), to adjust for differences in the reliability of estimates of QI rates from facilities of varying sizes, and to stabilize estimates for the smaller facilities, we used a Bayesian hierarchical model (i.e., a multivariate normal-binomial model) to calculate adjusted rates of each QI at each facility. Schwartz et al. (2012) have shown that MDS QI data are consistent with this model and that the adjusted rates calculated from the model are better predictors of the next year's observed rates than the current year's observed rates.

At each facility, we combined the 28 adjusted QI rates into a composite measure using facility-specific opportunity-based weights (Sullivan et al. 2013). This approach treats the impact of the adverse event measured by each quality indicator (QI) as equally important (Shwartz, Rosen, and Burgess 2017). The resulting composite can be thought of as the adjusted probability that an average resident experiences a QI event. We refer to this composite measure as the "quality score." We had monthly data on the MDS QI indicators from the third quarter of FY2009 to the third quarter of FY2012. To match the QI data to the Artifacts Tool data, we calculated the average of the quality score for the 3 months comprising each quarter.

Resource Utilization Groups (RUGs) Score. Facilities with more frail residents may find it more difficult to bring about RCC implementation, and independently of RCC implementation, more frail residents are more likely to experience adverse events. To measure facility case mix, we used the Resource Utilization Group (RUG)-III classification system. Based on the MDS data, residents are placed into one of a number of different categories defined by their expected resource use (Fries et al. 1994). Points are assigned to each category based on time studies documenting care needs for those in the category. The case mix for a facility is calculated by multiplying the number of patients in each RUGs category by the points for that category and summing the

results. VA uses the 53-RUG-III system (there are 53 categories) that went into effect in January 2006 for Medicare payment to skilled nursing facilities to calculate monthly RUGs scores. We calculated the RUGs score for the quarter by averaging the three monthly scores in the quarter.

Data Analysis

To examine whether more extensive RCC implementation is associated with higher quality over time, we ran a generalized estimating equation (GEE) model with quality score as the dependent variable and Artifacts Total Score as the independent variable of interest. Facility was treated as a fixed effect. This type of model can be thought of as examining the relationship between the quality score and Artifacts score over time within each facility and then aggregating the results across facilities.

We did not include a variable for time in the model for two reasons: (1) There was no overall trend in the quality score over time; and (2) the Artifacts score is a formal performance indicator within the VA, thus creating an explicit incentive to increase levels of RCC over time. As a result, there is a correlation between time and Artifacts score. We wanted any impact of the Artifacts score to be picked up by the Artifacts variable and not by a time variable correlated with that impact. We ran the GEE model with an identity link, a normal distribution (which is consistent with the assumption in the Bayesian model and is supported by the data; Shwartz et al. 2012) and an exchangeable correlation structure. As noted, there was not a great deal of variation in the quality measure over time, which in a fixed effects model creates a bias toward the null hypothesis (i.e., there is no relationship between the Artifacts score and the quality score). By running the GEE model with an exchangeable correlation rather than an autoregressive (1) correlation structure, we maximize the variation in the quality score that can be explained by the Artifacts score rather than attributing some of the variation to past values of the quality score. The RUGs score was included as a covariate in the initial GEE models. However, it was not significant in any of the models, and so we report results from the models excluding RUGs score.

Next, we examined the facility-specific trends in both the Artifacts score and quality score over time, and the relationship between the scores, by running separate individual regression models within each facility. For the trends analysis, time was the independent variable and each of the scores the dependent variable; for the relationship analysis, Artifacts score

was the independent variable and quality score the dependent variable. As our plan was to undertake a qualitative follow-up analysis to assess differences in characteristics of facilities that differ in patterns of change over time, we were as concerned about not identifying facilities that really were high or low performers as we were about incorrectly labeling an average facility as a high or low performer. Therefore, we used a more lenient p value of .10 as the threshold to identify statistically significant relationships. We analyzed the facility-specific trends by creating a 3×3 table. On one dimension, facilities were classified into “gets worse,” “no change,” and “gets better” based on the quality score; on the other dimension, they were similarly classified based on the Artifacts score. Within each cell, we show the number of facilities and the number for which there was a statistically significant (at the 0.10 level) positive relationship (improvement in Artifacts score is associated with higher quality) and a negative relationship (improvement in Artifacts score is associated with lower quality). To examine the relationship between the trend in scores and the baseline value of the scores, we converted both the Artifacts scores and the quality scores into z values [i.e., $(\text{score} - \text{mean score})/\text{SD score}$] and calculated the mean quality z score and the mean Artifacts score at baseline for the facilities shown in the margins of the table.

Finally, we examined characteristics of the facilities in which improvement in Artifacts score over time was associated with improved quality compared to those facilities in which improvement in Artifacts score over time was associated with lower quality. We compared geographic region, VA complexity level, mean patient days, mean RUGs score, mean adjusted costs per patient day, and baseline Artifacts score between the two groups, using chi-square tests for categorical variables and t -tests for continuous variables. VA complexity level is determined based on patient characteristics, clinical services offered, educational and research missions, and administrative complexity. Facilities are classified into three levels with Level 1 representing the most complex facilities, Level 2 moderately complex facilities, and Level 3 the least complex facilities. Costs per patient day were determined from the VA Managerial Cost Accounting (MCA) system, which combines expenditure and workload data with staff activity estimates to assign a cost to each intermediate product and then bundles these to patient-level costs. As nurse wages are adjusted to reflect variations in wages across geographic areas, we refer to the costs as adjusted costs. We calculated adjusted costs per patient day by dividing the adjusted cost by the average quarterly patient days.

RESULTS

Over the 13 quarters, the mean quality score of all facilities fluctuated only a little around the overall average of 0.14; the mean Artifacts score increased from 217 in FY2009 to 255.4 in FY2012 (see Figure 1). In the GEE model, the coefficient on the Artifacts score was positive, indicating higher Artifacts scores are associated with lower quality (as the quality score measures adverse events) but was not statistically significant ($p = .22$). As shown in Figure 1, at the time the new Artifacts Tool was introduced, there was an “interruption” in the monotonic increase of the Artifacts score over the quarters. We reran the GEE model with an indicator variable that was 0 for the first six quarters and 1 for the last seven. The Artifacts coefficients remained insignificant in that model. Finally, to evaluate the possibility that the impact of the new Artifacts Tool was more substantial than could be corrected by a rescoring of the items, we reran the model just using the last seven quarters of data. The Artifacts score coefficient in this model was positive and weakly significant ($p = .083$).

Figure 2 plots the coefficient for the Artifacts score when running a regression model with quality score as the dependent variable separately for each facility. Coefficients are plotted from lowest negative value (indicating higher RCC is associated with higher quality) to highest positive value (indicating higher RCC is associated with lower quality; statistically significant coefficients at the 0.10 level are indicated in red). For 36 facilities (of the 130), there was a statistically significant relationship between RCC and quality (well more than the 13 one would expect due to chance, using a p value of .10). For 15 facilities, higher RCC was associated with higher quality; for 21 facilities, higher RCC was associated with lower quality. If we had used a p value of .05,

Figure 1: Boxplots Displaying Quality Composite Scores and Artifacts of Culture Change Tool Scores over the Study Period [Color figure can be viewed at wileyonlinelibrary.com]

Figure 2: Artifacts Coefficients from Facility-Specific Models with Quality as the Dependent Variable and Artifacts as the Independent Variable. Black = $p < .05$; Gray = $p < .10$

there would have been eight facilities with a statistically significant negative coefficient and 13 with a positive coefficient.

Table 1 displays the distribution of sites by whether they improved, stayed the same, or got worse (using $p = .10$ threshold) in terms of RCC and quality over the study period. Over half of the facilities improved their RCC; at baseline, these facilities had lower RCC scores at baseline (0.27 SDs below the mean). Most of the other facilities did not have a statistically significant change in RCC; as a group, these facilities were slightly above average at baseline (0.20 SDs above the mean). The few facilities that got worse in terms of RCC were substantially above the mean at baseline (0.94 SDs). The relationship between baseline scores and change in quality was not as consistent; the facilities that got worse were in fact low at baseline (0.50 SDs below the mean). Cells within Table 1 show the number of cases in the cell in which there was a positive relationship between RCC and quality and the number with a

Table 1: Categorization of Sites by Changes over Time in RCC and Quality Scores

<i>Change in Quality</i>	<i>Descriptive Statistics</i>	<i>Change in RCC</i>			<i>Total CLCs Mean Baseline Quality</i>
		<i>Gets Worse</i>	<i>No Change</i>	<i>Gets Better</i>	
Gets better	<i>N</i>	0	10	9	19
	# CLCs + *		0	0	-.16
	# CLCs - *		0	7	
No change	<i>N</i>	5	32	39	76
	# CLCs + *	0	6	3	.27
	# CLCs - *	0	0	7	
Gets worse	<i>N</i>	2	14	19	35
	# CLCs + *	0	1	11	-.50
	# CLCs - *	1	0	0	
Total CLCs		7	56	67	130
Mean baseline RCC		.94	.20	-.27	

* $p < .10$; + indicates the coefficient in the facility-specific regression models is positive (higher RCC is associated with lower quality; - indicates the coefficient is negative (higher RCC is associated with higher quality)).

negative relationship. For example, of the 39 facilities that had a statistically significant improvement in RCC (at the 0.10 level) but no statistically significant change in quality, for seven there was a statistically significant positive relationship between RCC and quality; for three, there was a statistically significant negative relationship.

As shown in Table 2, mean adjusted cost per patient day was the only characteristic that differed between those facilities in which there was a positive relationship between RCC and quality and those in which there was a negative relationship. Mean adjusted costs (calculated as the average of mean adjusted cost per patient day of facilities in the group) were over 50 percent higher in the facilities with a negative relationship (\$167 vs. \$104).

DISCUSSION

Unlike a previous cross-sectional study that found higher levels of implementation of RCC were associated with higher quality (Sullivan et al. 2013), in this longitudinal study (which examines changes within facility over time) we did not find a statistically significant relationship between RCC implementation and quality of care as measured by the MDS indicators over the period

Table 2: Comparison of Available Facility Characteristics

	<i>Facilities Where Quality Improves and RCC Improves (n = 15)</i>	<i>Facilities Where Quality Decreases and RCC Improves (n = 21)</i>	<i>p Value</i>
<i>Categorical measures</i>			
Region			
East	9	8	.78
South	1	3	
Midwest	3	3	
West	2	7	
Complexity level			
1a	6	6	.78
1b	2	3	
1c	1	3	
2	1	3	
3	4	4	
<i>Continuous measures</i>			
Mean patient days	6,984.13	6,570.95	.80
Mean adjusted costs per patient days	103.91	167.10	.02
Mean RUGs	10.27	11.12	.31
Baseline artifacts total	225.40	212.14	.36

FY2009–FY2012. However, when we examined only the last seven quarters (the period after implementation of the new Artifacts Tool), we did find weak statistical evidence of a relationship ($p = .08$), suggesting the nature of the relationship was different in this period. Rather than RCC and quality being complementary, our analysis suggested a trade-off between RCC and quality; that is, greater RCC was associated with lower levels of quality. To some extent, this finding is not surprising. Extent of RCC implementation is a formal performance measure in the VA and factors into evaluation of CLC directors' performance. Although feedback on the quality measures is provided monthly to each CLC and CLCs are encouraged to use the measures to prioritize quality improvement efforts, CLC directors have not been held directly responsible for improvements beyond meeting accreditation. Thus, incentives encourage a focus on RCC at the expense of the quality measures. One might hypothesize that leaders focus on measures most important to their strategic priorities or those most easily improved. However, in our situation, one performance measure (RCC) is directly tied to financial rewards and the other measures (quality) only need to be good enough to avoid issues with accreditation. This situation is likely to change. A new set of performance metrics for CLC that include 11 of the MDS quality measures is likely to soon be

implemented. It will be interesting to analyze the impact of this change once the new measures are formally in place.

The relative importance of the RCC dimension compared to the quality dimension is apparent in Table 1. The 67 facilities that improved in terms of RCC had low scores at baseline, whereas the seven that declined had very high scores at baseline. Because extent of RCC implementation is a formal performance measure, there was strong motivation for the low-performing facilities to improve. Some of the high-performing facilities were able to devote their attention to other priorities, which probably accounts for their decline. The relationship between baseline scores and change in quality was not as consistent. The 19 facilities that got better over time had baseline scores slightly below the mean. But the 35 facilities that declined in terms of quality had baseline scores substantially below the mean. Because quality is not a formal performance measure, the incentive for these low performers to improve was substantially less than it was for the low RCC performers. This shows the power of incentivizing facility leaders: They will focus their efforts on those performance measures for which they are formally held responsible and deal with other measures as time, resources, and energy permits (Smith 1995; Goddard, Mannion, and Smith 2004; Franco-Santos, Lucianetti, and Bourne 2012; Johansson 2015).

Our findings are consistent with Holmstrom and Milgrom's theory of multitasking, which predicts that measuring and rewarding quality in some areas may harm quality in other areas (Holmstrom and Milgrom 1991). As noted by Werner, Konetzka, and Kruse, this is particularly true in cases where quality is multidimensional and quality improvement efforts target only some dimensions of quality (Werner, Konetzka, and Kruse 2009). In the broadest sense, one can consider both the RCC score and the MDS-based quality score as different dimensions of quality, one of which is incentivized and one of which is not. The theory predicts that providers will concentrate resources on the incentivized area. As Werner, Konetzka, and Kruse (2009) further note, if improvements in one dimension are structural (e.g., hiring more staff or better training staff), both dimensions of quality may improve. However, if the changes in one dimension are related only to that dimension (e.g., changes in particular processes) and resources are devoted to that dimension, it is more likely to see quality in the targeted dimension improve and quality in the other dimension decline. In our ongoing qualitative work (see below), we will examine the types of changes made in response to incentives to improve RCC to better understand the different patterns of changes in RCC score and the quality score.

We attempted to identify factors in our data that might distinguish facilities in which improvement in RCC was associated with higher quality from those where it was associated with lower quality. The only factor that was different between the two groups was adjusted cost per patient day. It may be that the lower costs are a result of the improvement in quality, as the events associated with the MDS quality measures are ones that might reasonably add to costs. We are currently analyzing in more depth the relationship between costs and quality over our study period.

Given these results, it would be useful to further study individual CLCs that improved on both RCC and quality to better understand their leadership, culture, processes, and practices that distinguished them from those that declined on quality. Grabowski et al. (2014) suggests the importance of qualitative work to understand the processes that lead to quality changes. For example, structural characteristics, organizational infrastructure, quality of care processes, and ways in which RCC was implemented are likely to be important factors distinguishing the high performers on both RCC and quality; these factors are best understood through qualitative methods. Our team is currently in the process of examining qualitative data from 12 sites—four sites high on both RCC/quality, four sites low on both RCC/quality, and four sites with mixed results on RCC/quality. The quantitative results from the current study were the basis for site selection. Our interview questions focus on RCC practices, clinical care processes, structural characteristics, and organizational infrastructure. We hope to elucidate factors distinguishing high- and low-performing facilities. Given the findings from our current study, we suggest that additional qualitative research is needed about monetary incentives and their impacts on quality. In addition, further research could focus on understanding whether residents are happier or have better quality of life in CLCs with higher RCC.

A major strength of this study is that we have longitudinal data from many CLCs collecting similar types of data. However, this study also has several limitations. The Artifacts of Culture Change Tool does not measure all domains of RCC. It may be the case that domains like staff–resident interactions and staff members' engagement of residents (Snow and Hartmann 2013) are more directly related to quality and then the dimensions measured by the Artifacts of Culture Change Tool. Future research could focus on other measures of quality such as the number and types of Joint Commission citations or other outcomes from Long Term Care Institute assessments.

ACKNOWLEDGMENTS

We would like to thank the project staff who assisted on the study, including Emily Rose Lopez, Omonyélé Adjognon, and Katelyn Gormley. We also wish to show our gratitude to our CHOIR colleagues for reviewing manuscripts and providing feedback (Amy Rosen, Hillary Mull, Marlena Shin, Qi Chen, Nathalie McIntosh, Barbara Lerner, and Chris Miller). The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the U.S. Department of Veterans Affairs.

Joint Acknowledgment/Disclosure Statement: This work was supported by VA HSR&D Grant IIR-11-356 (Sullivan, PI). The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the U.S. Department of Veterans Affairs.

Disclosures: None.

Disclaimer: None.

Previous dissemination: This work was presented at a 2015 Annual Gerontological Society of America meeting poster session (Orlando, FL).

REFERENCES

- Afendulis, C. C., D. J. Caudry, A. J. O'Malley, P. Kemper, D. C. Grabowski, and THRIVE Research Collaborative. 2016. "Green House Adoption and Nursing Home Quality." *Health Services Research* 51: 454–74.
- Franco-Santos, M., L. Lucianetti, and M. Bourne. 2012. "Contemporary Performance Measurement Systems: A Review of Their Consequences and a Framework for Research." *Management Accounting Research* 23 (2): 79–119.
- Fries, B. E., D. P. Schneider, W. J. Foley, M. Gavazzi, R. Burke, and E. Cornelius. 1994. "Refining a Case-mix Measure for Nursing Homes: Resource Utilization Groups (RUG-III)." *Medical Care* 32 (7): 668–85.
- Goddard, M., R. Mannion, and P. Smith. 2004. "Enhancing Performance in Health Care: A Theoretical Perspective on Agency and the Role of Information." *Studying the Organisation and Delivery of Health Services: A Reader* 9: 199.
- Grabowski, D. C., A. J. O'Malley, C. C. Afendulis, D. J. Caudry, A. Elliot, and S. Zimmerman. 2014. "Culture Change and Nursing Home Quality of Care." *The Gerontologist* 54 (suppl 1): S35–45.
- Grant, L. A., and E. McMahon. 2008. "The Cultural Evolution. Part One: Testing New Strategies to Drive Performance." *Provider* 34 (1): 35–6, 38–9.
- Hartmann, C. W., M. Shwartz, S. Zhao, J. A. Palmer, and D. R. Berlowitz. 2016. "Longitudinal Pressure Ulcer Rates Since the Adoption of Culture Change in

- Veterans Health Administration Nursing Homes." *Journal of the American Geriatrics Society* 64 (1): 151–5.
- Holmstrom, B., and P. Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, & Organization* 7: 24–52.
- Johansson, V. 2015. "When Will We Ever Learn?" *NISPAcee Journal of Public Administration and Policy* 8 (2): 149–70.
- Kane, R. A., T. Y. Lum, L. J. Cutler, H. B. Degenholtz, and T. C. Yu. 2007. "Resident Outcomes in Small-House Nursing Homes: A Longitudinal Evaluation of the Initial Green House Program." *Journal of the American Geriatrics Society* 55 (6): 832–9.
- Mueller, C. 2008. "Research in Culture Change in Nursing Homes." In *Nurses Involvement in Culture Change*. New York: Commonwealth Fund
- Port, C., P. D. Sloane, and S. Zimmerman. 2005. "Pragmatic Innovations in Long-Term Care: Developing a Research Agenda: Literature Review." *Paper prepared for conference Pragmatic Innovations in Long-Term Care*. Baltimore, MD. Available at <http://www.pragmaticinnovations.unc.edu/Pragmatic%20Innovations%20Literature%20Review.pdf>
- Rahman, A. N., and J. F. Schnelle. 2008. "The Nursing Home Culture-Change Movement: Recent Past, Present, and Future Directions for Research." *The Gerontologist* 48 (2): 142–8.
- Ronch, J. L. 2003. "Leading Culture Change in Long-Term Care: A Map for the Road Ahead." In *Culture Change in Long-Term Care*, edited by J. L. Ronch, and A. S. Weiner, pp. 65–80. Binghamton: Haworth Press.
- Shier, V., D. Khodyakov, L. W. Cohen, S. Zimmerman, and D. Saliba. 2014. "What Does the Evidence Really Say about Culture Change in Nursing Homes?" *The Gerontologist* 54 (suppl 1): S6–16.
- Shwartz, M., A. K. Rosen, and J. F. Burgess Jr. 2017. "Can Composite Measures Provide a Different Perspective on Provider Performance Than Individual Measures?" *Medical Care* 55 (12): e99–e103. <https://doi.org/10.1097/MLR.0000000000000407>
- Shwartz, M., E. A. Pekoz, C. L. Christiansen, J. F. Burgess Jr, and D. Berlowitz. 2012. "Shrinkage Estimators for a Composite Measure of Quality Conceptualized as a Formative Construct." *Health Services Research* 48 (1): 271–89.
- Smith, P. 1995. "On the Unintended Consequences of Publishing Performance Data in the Public Sector." *International Journal of Public Administration* 18 (2–3): 277–310.
- Snow, A., and C. W. Hartmann. 2013. "Increasing Engagement Between Staff and Residents: Using the RAISE Rating System as a Quality Improvement Tool." Paper presented at Pioneer Network Annual Meeting. Bellevue, WA.
- Stone, R., S. Reinhard, B. Bowers, D. Zimmerman, C. Phillips, C. Hawes, J. Fielding, and N. Jacobson. 2002. *Evaluation of the Wellspring Model for Improving Nursing Home Quality*. New York: Commonwealth Fund.
- Sullivan, J. L., M. Shwartz, J. F. J. Burgess, E. A. Peköz, C. L. Christiansen, M. Gerena-Melia, and D. Berlowitz. 2013. "Person-Centered Care Practices and Quality in Department of Veterans Affairs Nursing Homes: Is There a Relationship?" *Medical Care* 51 (2): 165–71.

- Werner, R. M., R. T. Konetzka, and G. B. Kruse. 2009. "Impact of Public Reporting on Unreported Quality of Care." *Health Services Research* 44 (2p1): 379–98.
- White-Chu, E. F., W. J. Graves, S. M. Godfrey, A. Bonner, and P. Sloane. 2009. "Beyond the Medical Model: The Culture Change Revolution in Long-Term Care." *Journal of the American Medical Directors Association* 10 (6): 370–8.

SUPPORTING INFORMATION

Additional supporting information may be found online in the supporting information tab for this article:

Appendix SA1: Author Matrix.