Nursing Facility Quality of Care and Costs:

Literature Review and Findings from a Special Analysis

Report Prepared for the Minnesota Department of Human Services in fulfillment of contract deliverables for the Evaluation of the Minnesota Nursing Facility Value-Based Reimbursement Initiative

By Greg Arling, Kathleen Abrahamson, Zachary Hass, and Dongjuan Xu

Purdue University School of Nursing

July 28 2020

Contents

Introduction

VBR Initiative

In 2015, the Minnesota legislature enacted major reforms to Medicaid nursing facility reimbursement. This new system is commonly referred to as "Value-Based Reimbursement" (VBR) was implemented on January 1, 2016. Goals of VBR were to:

- **Align Medicaid reimbursement rates more closely with the cost of caring for residents.**
- **Incentivize better care quality through the Medicaid rate setting process.**
- **Provide increased Medicaid reimbursement earmarked for direct care and care-related services.**
- **Improve efficiencies through other technical changes to Medicaid rate setting and payment.**

Main features of the VBR system are the application of a quality incentive payment for care-related services and a fixed price for other operating expenses. Nursing facility services are bundled into a comprehensive package of room, board and nursing services. Payment for this package of services is a daily per diem rate. The daily per diem rate can be further broken down into rate components of a care-related payment rate, other operating payment rate, external fixed costs payment rate, and a property rate.

Under VBR, care-related costs such as nurse wages and supplies, activities and social services are reimbursed at actual cost subject to a quality-based limit. Other operating costs such as housekeeping, laundry and property insurance are reimbursed using a pricing model, meaning the rate for these costs will be the same for all NFs in the state. The external fixed rate component is also established based on actual costs but is not subject to a limit. The property rate is determined through a facility-specific formula.

Evaluation Team

An evaluation team from the University of Minnesota and Purdue University has been conducting an ongoing external evaluation of VBR through a contract from the Minnesota Department of Human Services. In January 2019 the evaluation team prepared a major report, Evaluation of the NF Payment Reform Legislation: Background for the 2019 Report to the Legislature. The team will prepare a second major report in December 2020.

Issues Addressed in this Report

The current report provides a background for the December 2020 report. It contains five chapters addressing issues essential to the success of a value-based reimbursement system. Do we have effective measures of care quality? What is the relationship between costs, particularly expenditures for staffing and other care resources, and the quality of care? Chapter 1 contains a review of the research literature into the most widely used set of clinical quality measures, Medicare's Nursing Home Compare Quality Measures. It examines evidence about their reliability, validity, and association with other measures of care quality. Chapter 2 assesses the dimensionality and scoring of

Minnesota's nursing facility quality indicators, which are the state's version of clinical quality measures. In Chapter 3 reviews the research literature on the relationship between nursing facility expenditure and care quality. It examines evidence about the effectiveness of value-based reimbursement and other methods of incentivizing better care quality. The final two chapters present special analyses from Minnesota nursing facility data from 2013 to 2019. Chapter 4 applies a special statistical technique to cluster facilities having similar trajectories in their direct care costs and quality of care from before the VBR initiative (2013-2015) to after the initiative (2016-2019). It then compares the clusters according to facility characteristics, such as size, location, ownership type, and staffing levels. Finally, Chapter 5 delves into facility ownership changes from 2013 – 2019, to evaluate whether a change in ownership is associated with a decline in care quality.

Chapter 1

Study of the Nursing Home Quality Measures: A Literature Review

Dongjuan Xu, PhD MSN RN

Executive Summary

A search of the academic literature from 2010 to 2019 was completed to capture recent evidence surrounding nursing home quality measures. A total of 34 primary research articles from peer reviewed journals contributed to this report. Evidence fell into 3 broad categories: correlations between CMS 5-star quality ratings and ratings from other resources, relationships between nursing home quality measures and outcomes including quality of life, satisfaction, patient safety, negligence litigation, depression, urinary tract infection, and hospitalization or potentially preventable hospitalizations, and psychometric performance of individual quality measures.

The CMS 5-star ratings did not necessarily correlate with social media or online ratings. There was only minimal agreement on ranking of nursing homes between 5-star ratings and resident/family ratings. The 5-star ratings did not adequately reflect residents' quality of life, resident and family satisfaction, and resident safety. Certain quality ratings were related to negligence litigations, severity of depressive symptoms, or urinary tract infections; however, the effects were small. Available nursing home quality performance was not strongly or consistently associated with the risk of hospitalization or potentially preventable hospitalization. Only certain quality indicators appear to tap dimensions of clinical quality directly related to hospitalizations. Nursing homes with higher serious mental illness concentrations or higher proportion of African American residents were related to poorer care quality, while nursing homes with special care unit for dementia were related to better care quality. Moreover, the psychometric performance of individual quality measures was poor with low reliability and validity.

Search results are presented in the form of synthesized key findings, a summary of study findings organized by theme, and a table to provide an overview of individual studies.

Introduction

This report describes a search and review of academic literature that was completed to identify and summarize recent research regarding the correlations between different quality rating systems and between different raters, relationships between nursing home (NH) quality and outcomes, and psychometric performance of quality measures. The search addressed the following questions:

1) **What is correlation between different quality rating systems, and between different raters?**
2) **What is the relationship between NH quality and outcome variables such as quality of life, satisfaction, patient safety, negligence litigation, depression, urinary tract infection, and hospitalization?**
3) **What is psychometric performance of NH quality measures or quality indicators (QIs) such as relatability and validity?**

Search Methods

A search of PubMed was completed using the search terms (nursing home or nursing facility) AND (quality measure or quality indicator), resulting in 512 papers. Search limiters were published in English within the past 10 years since 2010 to highlight recent additions to the literature and peer-reviewed journals. Abstracts were reviewed and 24 usable papers were obtained. An additional 10 articles were identified through an ancestry search of the reference list of identified articles. A total of 34 primary research articles from peer reviewed journals contributed to this summary.

Search Results

Identified studies ranged in publication date from 2010-2019, and came from a wide variety of high quality nursing, gerontology, medical, economics and health services journals. Studies were most commonly retrospective analyses of large government databases such as the MDS, OSCAR, Medicare Claims Data, and the Area Resource File, with the exception of 1 qualitative study and 1 systematic review. Two studies conducted outside the United States with one study in Canada and 1 study in England.

Key Takeaways from the Synthesized Findings

1. Correlations between CMS 5-star ratings, and ratings from other sources

- **The CMS Nursing Home Compare (NHC) 5-star ratings did not necessarily correlate with social media or online ratings including Facebook, Yelp, and Google Consumer Reviews. Only one study found a moderate correlation.**
- **There was only minimal agreement on ranking of NHs between NHC 5-star ratings and resident/family ratings.**
- **There was a weak or no relationship between nursing home quality indicator performance and inspection deficiencies or complaints, although one study found 3 individual quality indicators (restraint use, worsened pressure ulcer, and experiencing worsened pain) predicted poor performance on inspections.**

2. Relationships between NH quality measures and different outcomes

- **The NHC 5-star ratings did not necessarily reflect nursing home residents' quality of life.**
- **The NHC 5-star ratings did not adequately reflect resident and family satisfaction.**
- **The relationships between NHC 5-star rating and resident safety measures were weak and inconsistent.**

- **There were significant associations between certain quality measures and negligence litigations; however, all the effects were small.**
- **The NHC 5-star ratings were not associated with new onset of depression; however, certain quality domain ratings were related to the severity of depressive symptoms.**
- **One study found that NHC 5-star ratings were significantly related to the occurrence of urinary tract infections.**
- **Available NH quality performance was not strongly or consistently associated with the risk of hospitalization or potentially preventable hospitalization. Only certain quality indicators appear to tap dimensions of clinical quality directly related to hospitalizations.**
- **Nursing homes with higher serious mental illness concentrations were related to poorer care quality.**
- **Nursing homes with special care unit for dementia were related to better care quality.**
- **Nursing homes with higher proportion of African American residents were related to worse care quality.**

3. Psychometric performance of quality measures

- **The agreements between falls recorded by facility staff in the MDS and falls recorded in the medical charts were fair or moderate.**
- **Depression QI was not a reliable and valid measure since it measured the ability of staff to detect depressive symptoms rather than the actual prevalence rate of depression.**
- **Incontinence QIs were not associated with clinically important differences in related care processes.**
- **Urinary tract infection QI overestimated the number of cases while adequately screening out residents without infections.**
- **Weight loss QI was able to discriminate differences in prevalence of weight loss between facilities.**
- **Restraint QI was not able to discriminate differences in the use of restraining devices.**
- **Pressure ulcer QI was not an effective measure of the quality of pressure ulcer care in facilities and the QI score could be misleading.**
- **There was potential systematic bias in reporting pain QI.**
- **Three short-stay quality measures (rehospitalizations, ED visits, and successful discharges to the community) were weakly correlated.**

Summary of Study Findings

1. CMS Nursing Home Compare 5-star ratings vs. ratings from other sources (8 studies)

1.1 social media ratings (3 studies with inconsistent findings: 1 moderate correlation; 1 weak correlation; and 1 no correlation)

Li et al. (2019) studied the correlations between NHC 5-star overall ratings, experience-of-care survey ratings from family members or legal guardians/representatives, and average score of 5-star ratings from 4 popular social media or online review sites (Facebook, Yelp, Google Consumer Reviews, and Caring.com) among 196 NHs in Maryland from July 2015 to July 2017 (Li, Cai, & Wang, 2019). They found the moderate correlation (ρ=0.41, $p<0.001$) between social media rating and NHC overall rating, moderate-to-strong correlations (ρ ranged from 0.40-0.57, $p<0.001$) between social media rating and experience-of-care ratings, and moderate correlations (ρ about 0.35, $p<0.001$) between social media rating and individual NHC quality measures including number of deficiency citations, adjusted RN staffing, adjusted total nurse staffing and number of complaint. The NHC overall rating was

moderately to strongly correlated with experience-of-care ratings (ρ ranged from 0.30-0.53, p<0.001), and moderately correlated with individual NHC quality measures (ρ ranged from 0.39-0.49, p<0.001). Johari et al. (2018) compared Yelp ratings with NHC ratings among 675 NHs in California between September and November 2016 and found weak correlations (Johari et al., 2018). Hefele et al. (2018) compared Facebook user ratings with NHC ratings and resident/family satisfaction/experience survey ratings among 35 NHs in Maryland and 78 NHs in Minnesota. They found Facebook ratings were not significantly correlated with NHC ratings or survey-based resident/family satisfaction ratings (Hefele, Li, Campbell, Barooah, & Wang, 2018).

1.2 resident/family ratings (3 studies: 2 studies with minimal agreement on ranking of NHs and 1 qualitative study)

Mukamel et al (2016) asked residents and family members to use the Nursing Home Compare Plus app to create their own composite quality scores based on their own preferences and medical needs among 146 patients who were discharged from the hospital to NHs (Mukamel et al., 2016). They found that residents differed from each other and from CMS in the number of quality measures they chose to include in their composite and in their weighting of each quality measure. Moreover, there was only minimal agreement on ranking of NHs (Kappa statistics ranged from 0.22-0.38) between NHC 5-star ratings and resident/family ratings (Mukamel et al., 2016). Çalıkoğlu et al. (2012) also found very low agreement (Kappa statistics ranged from 0.15 for health inspections to 0.04 for quality measures) between NHC 5-star ratings and the ranking based on family experience survey results among 208 NHs in Maryland between September and December 2009. Moreover, family ratings of experience of care were significantly related to two 5-star domains (health inspections and nursing staff), but not to the quality domain (Çalıkoğlu, Christmyer, & Kozlowski, 2012).

Schapira et al. (2016) conducted a qualitative study among 35 persons or family members recently admitted or anticipating admission to a NH in Philadelphia from October 2013 to August 2014 and explored their responses to both content and format of the NHC report card. Although star ratings, clinical quality measures, and benchmarking information were salient to their decision-making process, participants had confusions about 1) the mathematical relationship between the overall rating composite score and the three components: health inspection, staffing, and quality measure ratings (participants expected the overall rating to be an average of the three domain-specific ratings.), and 2) the inconsistent direction of the quality measures rates (high quality could be indicated by lower values or higher values) (Schapira, Shea, Duey, Kleiman, & Werner, 2016).

1.3 quality indicators and inspection performance or NH complaints (2 studies: little relationships between quality indicators and inspection performance or complaints)

Mashouri et al. (2019) investigated whether quality indicators (2016-2017) could predict future inspection performance (2017-2018, three classes: in good standing, needing improvement, and needing significant improvement) in 594 LTC facilities in Ontario, Canada and found a weak relationship with a classification accuracy of 40.1%. They also found only 3 individual quality indicators (restraint use, worsened pressure ulcer, and experiencing worsened pain) predicted poor performance on inspections and one quality indicator (improved physical functioning) had a unexpected, inverse relationship with LTC facilities predicted as being in good standing (Mashouri, Taati, Quirt, & Iaboni, 2019).

Troyer and Sause (2013) examined the associations between 4 quality indicators (incontinence without a toileting plan, indwelling catheters, decline in late-loss ADLs, and pressure sores among high-risk residents) and two sources of resident- and caregiver-derived NH complaints (North Carolina Long-Term Care Ombudsman Program and state certification agency) from 2002-2006. They found

that quality indicators were unrelated to the volume of both types of complaints, and inspection deficiencies were positively associated with state certification agency complaints (Troyer & Sause, 2013).

2. Quality measures and different outcomes (14 studies)

2.1 quality of life (2 studies: star rating did not reflect quality of life)

Netten et al. (2012) found no significant relationship between NH star rating and social care-related quality of life after controlling for resident and NH characteristics in England (Netten et al., 2012). Similarly, Kim et al. (2014) found no significant relationship between NHC's overall 5-star quality rating and quality of life (p=0.12) among 251 long-stay NH residents with preserved cognition in 32 NHs in Detroit (S. J. Kim et al., 2014).

2.2 satisfaction (one study: star rating did not reflect consumer satisfaction)

Williams et al. (2016) indicated that 5-star NHs had significantly higher satisfaction score than 1-star NHs, however, there were inconsistencies in the categorical comparisons between the NHC overall ratings and the consumer satisfaction categories in 918 NHs in Ohio. Many 5-star NHs had moderate to very low consumer satisfaction (54% compared with resident satisfaction and 41% compared with family satisfaction), and many 1-star NHs had high to very high consumer satisfaction (19% compared with resident satisfaction and 20.0% compared with family satisfaction). The findings indicate that NHC 5-star rating system does not adequately reflect resident and family satisfaction (Williams, Straker, & Applebaum, 2016).

2.3 patient safety (one study: star rating did not reflect patient safety)

Brauner et al. (2018) found the relationship between NHC 5-star rating and six measures of resident safety (injurious falls, urinary tract infections, pressure sores among long-stay residents, pressure sores among short-stay residents, and two measures of medication errors) was weak and somewhat inconsistent (ρ ranged from 0.05-0.21) in 15652 NHs in the first quarter of 2017. Although 1-star NHs had higher rates of adverse safety events and 5-star NHs had the lowest rates, for NHs with two, three, or four stars, there was no meaningful difference in adverse safety events (Brauner et al., 2018).

2.4 negligence litigation (two studies: weak or small associations)

Studdert et al. (2011) examined whether high-quality NHs were less likely to be sued for negligence among 1465 NHs between 1998 and 2006, and found that NHs with more deficiencies (OR=1.09), with more serious deficiencies (OR=1.04), having more residents with weight loss (OR=1.05), and with pressure ulcers (OR=1.09), had higher odds of being sued; however, all these effects were relatively small (Studdert, Spittal, Mello, O'Malley, & Stevenson, 2011). Stevenson et al. (2013) investigated whether the experience of being sued and incurring litigation costs impacted the quality of care subsequently delivered in 1514 NHs between 1998 and 2010, and found that higher litigation costs were related to lower subsequent quality and only four of the 27 examined associations were statistically significant with small effect sizes (Stevenson, Spittal, & Studdert, 2013).

2.5 depression (one study: significant association with severity of depressive symptoms, not with new onset of depression)

Yuan et al. (2019) examined the association between NHC star ratings and new onset of depression and severity of depressive symptoms at 90 days in 129837 long-stay residents without indictors of depression admitted to 13921 NHs. They found that star quality ratings was not associated with new

onset of depression and lower quality domain ratings were related to more severe depressive symptoms (Yuan, Lapane, Baek, Jesdale, & Ulbricht, 2019).

2.6 urinary tract infection (one study with significant association with UTI)

Gucwa et al. (2016) investigated the association between NHC 5-star quality rating and urinary tract infection among 1523 residents in 12 skilled nursing facilities in Long island and found that overall star ratings were significantly related to the occurrence of urinary tract infections (Gucwa, Dolar, Ye, & Epstein, 2016). The interpretation of findings needs to be cautious because there were only 12 facilities and overall quality ratings and three domain ratings (health inspection, nursing staff, and quality) were included in the models at the same time.

2.7 hospitalization or potentially preventable hospitalization (6 studies: no or weak associations)

Using 2003 to 2006 Medicaid data in Florida, a study found that a lack of association between quality deficiencies and time to first potentially preventable hospitalization (PPH) (Becker, Boaz, Andel, Gum, & Papadopoulos, 2010). A study used 1998 to 2004 MDS state data in New York to show that facility-level deficiencies were associated with a decreased time to first hospitalization or the time between subsequent hospitalizations (O'Malley, Caudry, & Grabowski, 2011). A study using national Medicare data on fee-for-service Medicare beneficiaries discharged to a skilled nursing facility after an acute care hospitalization between 2009 and 2010, indicated that quality deficiency rating (5 star vs. 1 star; the higher star means less deficiencies) and the proportion of post-acute care residents with new or worsening pressure ulcers (25th percentile vs. 75th percentile) were negatively associated with 30 days hospital readmission and death, respectively (Neuman, Wirtalla, & Werner, 2014). However, the difference was very small (Neuman et al., 2014). Using national data of long-stay NH residents, a study indicated that the NHC star rating had weak correlations with rates of PPH and potentially preventable ED visits (Fuller, Goldfield, Hughes, & McCullough, 2019). Xu et al. (2019) found that available quality indicators were not strongly or consistently associated with the risk of hospitalization (neither overall nor PPH). Among these 23 quality indicators, 5 quality indicators (antipsychotics without a diagnosis of psychosis, unexplained weight loss, pressures sores, balder continence, and ADL dependence) were related significantly to hospitalization and only 4 quality indicators (antipsychotics without a diagnosis of psychosis, unexplained weight loss, ADL dependence, and urinary tract infections) were related to PPH (Xu, Kane, & Arling, 2019) . Snyder et al. (2019) found that NHC overall rating and two domain ratings (health inspection and quality measure) were not associated with 90-day readmission/major complications, >75th percentile post-acute cost, and 90-day bundle cost exceeding the target price among 488 patients who discharged to 105 skilled nursing facilities after primary total joint arthroplasty. The higher level of nursing staff domain rating was significantly associated with a decreased odds of the two cost outcomes (Snyder et al., 2019).

3. Nursing homes with higher prevalence of serious mental illness or dementia (7 studies)

3.1 NHs with higher serious mental illness (SMI) concentrations (3 studies: associated with poorer quality)

Kim et al. (2013) examined the prevalence of SMI and three mental health-related quality measures (depression without antidepressant therapy, bladder/bowel incontinence without a toileting plan, and the use of physical restraint in residents with dementia) among 135 Veterans Affairs (VA) NHs between fiscal years 2005-2007. They found that NHs with higher prevalence of SMI was associated with poorer quality of the three measures after adjusting for time and other facility-level characteristics (H. M. Kim et al., 2013). In addition, the presence of special care unit for dementia was associated with higher odds of physical restraint use (H. M. Kim et al., 2013). McGarry et al. (2019) found similar results that admission to NHs with high concentration of residents with SMI (at least

10% of a facility's proportion having an SMI diagnosis) was related to poorer quality for both residents with and without SMI among 58571 residents in 12027 NHs from 2006-2010. Particularly, relative to residents admitted to a low-SMI facility, for residents with SMI, admission to a high-SMI facility was associated with a 3.7 percentage point increase in the probability of feeding tube use; for residents without SMI, admission to a high-SMI facility was associated with higher probability of catheter use (a 1.7 percentage point increase), being hospitalized (a 3.8 percentage point increase), and having a feeding tube (a 2.1 percentage point increase) (McGarry et al., 2019).Rahman et al. (2013) found that NHs with an increase in the share of SMI was related to higher rates of hospitalization for residents without SMI and lower staffing skill mix and level (ratio of RN to total nurses and direct care hours per resident day)(Rahman, Grabowski, Intrator, Cai, & Mor, 2013).

3.2 presence of special care unit (SCU) for dementia (3 studies)

Joyce et al. (2018) found that NHs with an SCU was associated with a decrease in inappropriate antipsychotics (-9.7 percent), physical restraints (-9.6 percent), pressure ulcers (-3.3 percent), feeding tubes (-8.3 percent), and hospitalizations (-14.7 percent) among 704782 residents with dementia. They found no association with the use of indwelling urinary catheters (Joyce, McGuire, Bartels, Mitchell, & Grabowski, 2018).

Nazir et al. (2011) found that the incidence rates of the worsening behavior QI was significantly higher for residents on SCUs than for residents on conventional unit, and the worsening behavior QI scores after adjusted for cognitive impairment and presence of SCU produced significant shifts in NH rankings, which providing fairer comparison for NHs to take care of residents with dementia (Nazir, Arling, Perkins, & Boustani, 2011). Nazir et al. (2012) found similar results regarding the prevalence of falls QI among 21587 residents in 381 NHs, that is, the prevalence of falls was significantly higher for residents on SCUs than for residents on conventional unit, and there was a non-linear and significant association between the prevalence of falls and residents' level of cognitive impairment. The fall QI adjusted for cognitive impairment and presence of SCU provided a more accurate measure of NH care (Nazir, Mueller, Perkins, & Arling, 2012).

3.3 NHs with higher proportion of African American residents (one study)

Rivera-Hernandez et al. (2019) found that skilled nursing facilities with higher proportions of African American residents had worse quality of post-acute care as measured by 30-day rehospitalization rate, successful discharge from the facility to the community, and five-star quality ratings among 649187 Medicare beneficiaries from 8375 facilities(Rivera-Hernandez, Rahman, Mukamel, Mor, & Trivedi, 2019).

4. Reliability and validity of quality measures (4 studies, one of them is a systematic review)

Hutchinson et al. (2010) conducted a systematic review about the reliability and validity of RAI-MDS 2.0 quality indicators (14 articles and 1 report were included) and concluded that the strength of the evidence was limited (Hutchinson et al., 2010). Regarding *falls* QI, the agreements between falls recorded by facility staff in the MDS and falls recorded in the medical charts were fair (Kappa statistic = 0.29) for a 30-day timeframe and moderate (Kappa statistic = 0.29) for a 180-day timeframe, and MDS underreported falls (Hill-Westmoreland & Gruber-Baldini, 2005). Regarding *depression* QI, all included 3 studies suggested depression QI was not a reliable and valid measure: it measured the ability of staff to detect depressive symptoms rather than the actual prevalence rate of depression (Schnelle et al., 2001), it should not be interpreted as discriminating either differential rates of depression or care quality in relation to depression (Simmons et al., 2004), and it correlated poorly with the valid instruments and exhibited inferior sensitivity and specificity (Heiser, 2004). Regarding *incontinence* QIs, they were not associated with clinically important differences in related care

processes (Schnelle et al., 2003). Regarding *urinary tract infection* QI, it overestimated the number of cases (only 13.9% could be validated as correct) while adequately screening out residents without infections (98.2% of residents without experiencing UTI could be validated as correct) (Stevenson et al., 2004). Regarding *weight loss* QI, it may have concurrent validity since it was able to discriminate differences in prevalence of weight loss between facilities (Simmons et al., 2003). Regarding *restraint* QI, it was not able to discriminate differences in the use of restraining devices when the resident was out of bed (Schnelle et al., 2004). Regarding *pressure ulcers* QI, it was not an effective measure of the quality of pressure ulcer care in facilities and the QI score could be misleading (Bates-Jensen et al., 2003). Regarding *pain* QI, high prevalence scores were associated with more frequent pain assessment and appropriate pain-related care practices, as opposed to poor care quality (Cadogan et al., 2004), and there was potential systematic bias in reporting pain QI (Roy & Mor, 2005).

Estabrooks et al. (2013) used modified Delphi technique and asked 16 experts to rank a list of 13 MDS 2.0 quality indicators based on practice sensitivity. In the top 5, pressure ulcers were the most practice sensitive QI, followed by worsening pain, physical restraint use, the use of antipsychotic medications without a diagnosis of psychosis, and indwelling catheters (Estabrooks, Knopp-Sihota, & Norton, 2013).

Werner et al. (2013) explored the link between processes and outcomes of care and tested the extent to which improvements in outcomes of care were explained by changes in nursing home processes among 16,623 NHs from 2000 to 2009. They found that of the 5 outcome quality measures examined (pain, incontinence, pressure scores, and weight loss), only improvements in the percentage of long-stay NH residents in moderate or severe pain were associated with changes in NH processes of care with very small effect (Werner, Konetzka, & Kim, 2013).

Saliba et al. (2018) found that the three short-stay quality measures (rehospitalizations, ED visits, and successful discharges to the community) were weakly correlated. The correlation between the ED and the rehospitalizations measures was 0.25, between discharge to community and rehospitalization was −0.3, and between ED visits and discharge to the community was −0.05 (Saliba, Weimer, Shi, & Mukamel, 2018).

5. A composite measure of quality (one study)

Shwartz et al. (2013) compared composite scores calculated from the 28 QIs using both observed rates and shrunken rates derived from a Bayesian multivariate normal-binomial model in 112 Veterans Health Administration NHs in fiscal years 2005–2008. They found that shrunken-rate composite scores in 1 year had better prediction of the observed total number of QI events or the observed-rate composite scores in the following year (Shwartz, Peköz, Christiansen, Burgess Jr, & Berlowitz, 2013).

References

Becker, M. A., Boaz, T. L., Andel, R., Gum, A. M., & Papadopoulos, A. S. (2010). Predictors of preventable nursing home hospitalizations: the role of mental disorders and dementia. *The American Journal of Geriatric Psychiatry, 18*(6), 475-482.

Brauner, D., Werner, R. M., Shippee, T. P., Cursio, J., Sharma, H., & Konetzka, R. T. (2018). Does nursing home compare reflect patient safety in nursing homes? *Health Affairs, 37*(11), 1770-1778.

Çalıkoğlu, Ş., Christmyer, C. S., & Kozlowski, B. U. (2012). My Eyes, Your Eyes—The Relationship between CMS Five-Star Rating of Nursing Homes and Family Rating of Experience of Care in Maryland. *Journal for Healthcare Quality, 34*(6), 5-12.

Estabrooks, C. A., Knopp-Sihota, J. A., & Norton, P. G. (2013). Practice sensitive quality indicators in RAI-MDS 2.0 nursing home data. *BMC research notes, 6*(1), 460.

Fuller, R. L., Goldfield, N. I., Hughes, J. S., & McCullough, E. C. (2019). Nursing home compare star rankings and the variation in potentially preventable emergency department visits and hospital admissions. *Population health management, 22*(2), 144-152.

Gucwa, A. L., Dolar, V., Ye, C., & Epstein, S. (2016). Correlations between quality ratings of skilled nursing facilities and multidrug-resistant urinary tract infections. *American journal of infection control, 44*(11), 1256-1260.

Hefele, J. G., Li, Y., Campbell, L., Barooah, A., & Wang, J. (2018). Nursing home Facebook reviews: who has them, and how do they relate to other measures of quality and experience? *BMJ Qual Saf, 27*(2), 130-139.

Hutchinson, A. M., Milke, D. L., Maisey, S., Johnson, C., Squires, J. E., Teare, G., & Estabrooks, C. A. (2010). The resident assessment instrument-minimum data set 2.0 quality indicators: a systematic review. *BMC health services research, 10*(1), 166.

Johari, K., Kellogg, C., Vazquez, K., Irvine, K., Rahman, A., & Enguidanos, S. (2018). Ratings game: an analysis of nursing home compare and Yelp ratings. *BMJ quality & safety, 27*(8), 619-624.

Joyce, N. R., McGuire, T. G., Bartels, S. J., Mitchell, S. L., & Grabowski, D. C. (2018). The impact of dementia special care units on quality of care: An instrumental variables analysis. *Health services research, 53*(5), 3657-3679.

Kim, H. M., Banaszak-Holl, J., Kales, H., Mach, J., Blow, F., & McCarthy, J. F. (2013). Trends and predictors of quality of care in VA nursing homes related to serious mental illness. *Medical Care*, 659-665.

Kim, S. J., Park, E.-c., Kim, S., Nakagawa, S., Lung, J., Choi, J. B., . . . Kim, K. (2014). The association between quality of care and quality of life in long-stay nursing home residents with preserved cognition. *Journal of the American Medical Directors Association, 15*(3), 220-225.

Li, Y., Cai, X., & Wang, M. (2019). Social media ratings of nursing homes associated with experience of care and "Nursing Home Compare" quality measures. *BMC health services research, 19*(1), 260.

Mashouri, P., Taati, B., Quirt, H., & Iaboni, A. (2019). Quality Indicators as Predictors of Future Inspection Performance in Ontario Nursing Homes. *Journal of the American Medical Directors Association*.

McGarry, B. E., Joyce, N. R., McGuire, T. G., Mitchell, S. L., Bartels, S. J., & Grabowski, D. C. (2019). Association between High Proportions of Seriously Mentally Ill Nursing Home Residents and the Quality of Resident Care. *Journal of the American Geriatrics Society, 67*(11), 2346-2352.

Mukamel, D. B., Amin, A., Weimer, D. L., Sharit, J., Ladd, H., & Sorkin, D. H. (2016). When patients customize nursing home ratings, choices and rankings differ from the government's version. *Health Affairs, 35*(4), 714-719.

Nazir, A., Arling, G., Perkins, A. J., & Boustani, M. (2011). Monitoring quality of care for nursing home residents with behavioral and psychological symptoms related to dementia. *Journal of the American Medical Directors Association, 12*(9), 660-667.

Nazir, A., Mueller, C., Perkins, A., & Arling, G. (2012). Falls and nursing home residents with cognitive impairment: new insights into quality measures and interventions. *Journal of the American Medical Directors Association, 13*(9), 819. e811-819. e816.

Netten, A., Trukeschitz, B., Beadle-Brown, J., Forder, J., Towers, A.-M., & Welch, E. (2012). Quality of life outcomes for residents and quality ratings of care homes: is there a relationship? *Age and ageing, 41*(4), 512-517.

Neuman, M. D., Wirtalla, C., & Werner, R. M. (2014). Association between skilled nursing facility quality indicators and hospital readmissions. *JAMA, 312*(15), 1542-1551.

O'Malley, A. J., Caudry, D. J., & Grabowski, D. C. (2011). Predictors of nursing home residents' time to hospitalization. *Health services research, 46*(1p1), 82-104.

Rahman, M., Grabowski, D. C., Intrator, O., Cai, S., & Mor, V. (2013). Serious mental illness and nursing home quality of care. *Health services research, 48*(4), 1279-1298.

Rivera-Hernandez, M., Rahman, M., Mukamel, D. B., Mor, V., & Trivedi, A. N. (2019). Quality of post-acute care in skilled nursing facilities that disproportionately serve black and Hispanic patients. *The Journals of Gerontology: Series A, 74*(5), 689-697.

Saliba, D., Weimer, D. L., Shi, Y., & Mukamel, D. B. (2018). Examination of the new short-stay nursing home quality measures: rehospitalizations, emergency department visits, and successful returns to the community. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing, 55*, 0046958018786816.

Schapira, M. M., Shea, J. A., Duey, K. A., Kleiman, C., & Werner, R. M. (2016). The nursing home compare report card: perceptions of residents and caregivers regarding quality ratings and nursing home choice. *Health services research, 51*, 1212-1228.

Shwartz, M., Peköz, E. A., Christiansen, C. L., Burgess Jr, J. F., & Berlowitz, D. (2013). Shrinkage estimators for a composite measure of quality conceptualized as a formative construct. *Health services research, 48*(1), 271-289.

Snyder, D. J., Kroshus, T. R., Keswani, A., Garden, E. B., Koenig, K. M., Bozic, K. J., . . . Moucha, C. S. (2019). Are Medicare's Nursing Home Compare Ratings Accurate Predictors of 90-Day Complications, Readmission, and Bundle Cost for Patients Undergoing Primary Total Joint Arthroplasty? *The Journal of arthroplasty, 34*(4), 613-618.

Stevenson, D. G., Spittal, M. J., & Studdert, D. M. (2013). Does litigation increase or decrease health care quality? A national study of negligence claims against nursing homes. *Medical Care, 51*(5), 430.

Studdert, D. M., Spittal, M. J., Mello, M. M., O'Malley, A. J., & Stevenson, D. G. (2011). Relationship between quality of care and negligence litigation in nursing homes. *New England Journal of Medicine, 364*(13), 1243-1250.

Troyer, J. L., & Sause, W. (2013). Association between traditional nursing home quality measures and two sources of nursing home complaints. *Health services research, 48*(4), 1256.

Werner, R. M., Konetzka, R. T., & Kim, M. M. (2013). Quality improvement under nursing home compare: the association between changes in process and outcome measures. *Medical Care, 51*(7), 582.

Williams, A., Straker, J. K., & Applebaum, R. (2016). The nursing home five star rating: How does it compare to resident and family views of care? *The Gerontologist, 56*(2), 234-242.

Xu, D., Kane, R., & Arling, G. (2019). Relationship between nursing home quality indicators and potentially preventable hospitalisation. *BMJ quality & safety, 28*(7), 524-533.

Yuan, Y., Lapane, K. L., Baek, J., Jesdale, B. M., & Ulbricht, C. M. (2019). Nursing Home Star Ratings and New Onset of Depression in Long-Stay Nursing Home Residents. *Journal of the American Medical Directors Association, 20*(10), 1335-1339. e1310.

Table 1. Summary of Studies

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
Becker, M. A., Boaz, T. L., Andel, R., Gum, A. M., & Papadopoulos, A. S. (2010). Predictors of preventable nursing home hospitalizations: the role of mental disorders and dementia. The American Journal of Geriatric Psychiatry, 18(6), 475-482.	To examine risk factors for hospitalization of Medicaid-enrolled NH residents with "ambulatory care-sensitive" conditions	retrospective secondary data analyses	72,251 Medicaid-enrolled NH residents in 647 NHs in Florida during fiscal year 2003–2006	quality measures and hospitalization	Residents from for-profit facilities, facilities that were not a member of a chain, had more Medicaid recipients, and fewer than 120 beds had greater risk of "ambulatory care-sensitive" hospitalizations. There was no association between quality deficiencies and potentially preventable hospitalization.	single state	Attention to the identified predictors of hospitalization for "ambulatory care-sensitive" conditions, which are potentially preventable, could reduce the risk and cost of these hospitalizations among Medicaid-enrolled NH residents.
Brauner, D., Werner, R. M., Shippee, T. P., Cursio, J., Sharma, H., & Konetzka, R. T. (2018). Does nursing home compare reflect patient safety in nursing homes? Health Affairs, 37(11), 1770-1778.	To compare NHs' performance on several composite quality measures from NHC to their performance on measures of patient safety in NHs such as	retrospective secondary data analyses	15652 NHs in the first quarter of 2017	quality measures and patient safety	Although NHC captures some aspects of patient safety, the relationship was weak and somewhat inconsistent, leaving consumers who care about patient safety with little guidance.	national data	NHC should be refined to provide a clearer picture of patient safety and quality of life, allowing consumers to weight these domains

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
	pressure sores, infections, falls, and medication errors.						according to their preferences and priorities.
Çalıkoğlu, Ş., Christmyer, C. S., & Kozlowski, B. U. (2012). My Eyes, Your Eyes—The Relationship between CMS Five-Star Rating of Nursing Homes and Family Rating of Experience of Care in Maryland. Journal for Healthcare Quality, 34(6), 5-12.	To look at the relationship between NHC QMs obtained from assessments conducted by NH staff and the family rating of overall care using data from Maryland NHs.	retrospective secondary data analyses	208 NHs in Maryland between September and December 2009	resident/family ratings	strong positive correlation between family experience of care score and two five-star domains, namely health inspections and nurse staffing, and no relationship with the quality domain	single state	The lack of relationship between the quality domain and the family score may be due to inadequate risk adjustment or that each rating system measures different aspects of quality
Estabrooks, C. A., Knopp-Sihota, J. A., & Norton, P. G. (2013). Practice sensitive quality indicators in RAI-MDS 2.0 nursing home data. BMC research notes, 6(1), 460.	to identify practice sensitive QIs which believed to be the most sensitive to clinical practice.	qualitative study	16 experts to rank a list of 13 MDS 2.0 quality indicators based on practice sensitivity	validity and reliability of the QIs	Pressure ulcers were identified as the most practice sensitive QI followed by worsening pain, physical restraint use, the use of antipsychotic medications without a diagnosis of psychosis, and indwelling catheters. When	qualitative study	Focusing on these 13 practice sensitive QIs provides both the greatest potential for improving resident function and slowing the

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
					stratified by informant group, although the top five QIs stayed the same, the ranking of the 13 QIs differed by group		trajectory of decline that most residents experience.
Fuller, R. L., Goldfield, N. I., Hughes, J. S., & McCullough, E. C. (2019). Nursing home compare star rankings and the variation in potentially preventable emergency department visits and hospital admissions. Population health management, 22(2), 144-152.	To examine the NHC Stars measure and the rates of potentially preventable hospital admissions and potentially preventable ED visits	retrospective secondary data analyses	a subset of 439,011 long-term residents residing in 12,883 NHs from 2010–2011	quality measures and hospitalization	the NHC Stars measure has limited correlation with rates of the potentially preventable hospital admissions and potentially preventable ED visits	national data	weak correlations
Gucwa, A. L., Dolar, V., Ye, C., & Epstein, S. (2016). Correlations between quality ratings of skilled nursing facilities and multidrug-resistant urinary tract infections. American	To determine risk factors for the acquisition of urinary tract infections and multidrug-resistant organisms in residents of	primary and secondary data analyses	1523 residents in 12 skilled nursing facilities in Long island	quality measures and urinary tract infections	Overall quality rating predicted the occurrence of urinary tract infections, whereas identification of multidrug-resistant organisms was dependent on the level of nursing care	There were only 12 facilities and overall quality ratings and three domain ratings	The CMS's quality rating system may provide some insight into the status of infection control practices in SNFs.

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
journal of infection control, 44(11), 1256-1260.	skilled nursing facilities				received. The mean predicted probability of urinary tract infections and receipt of contaminated samples was inversely dependent on the facility's rating, where the likelihood increased as overall quality ratings decreased.	(health inspection, nursing staff, and quality) were included in the models at the same time.	
Hefele, J. G., Li, Y., Campbell, L., Barooah, A., & Wang, J. (2018). Nursing home Facebook reviews: who has them, and how do they relate to other measures of quality and experience? BMJ Qual Saf, 27(2), 130-139.	To examine the relationship between Facebook user-generated NH ratings and other measures of NH satisfaction/experience and quality	retrospective secondary data analyses	35 NHs in Maryland and 78 NHs in Minnesota	NHC 5-star ratings vs. ratings from other sources	Facebook ratings were not significantly correlated with the CMS 5-star rating or survey-based resident/family satisfaction ratings.	two states	Given the disconnect between Facebook ratings and other, more scientifically grounded measures of quality, concerns about the validity and use of social media ratings are warranted.
Hutchinson, A. M., Milke, D. L., Maisey, S., Johnson, C., Squires, J. E., Teare,	to systematically examine published and grey research	systematic review	14 articles and one report examining	validity and reliability of the QIs	Studies about 10 QIs including falls, depression, depression without treatment,	systematic review	Evidence for the reliability and validity of the QIs remains

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
G., & Estabrooks, C. A. (2010). The resident assessment instrument-minimum data set 2.0 quality indicators: a systematic review. BMC health services research, 10(1), 166.	reports in order to assess the state of the science regarding the validity and reliability of the RAI-MDS 2.0 QIs		the validity and/or reliability of the RAI-MDS 2.0 QIs were included. All studies were conducted in the US and included from one to a total of 209 facilities.		urinary incontinence, urinary tract infections, weight loss, bedfast, restraint, pressure ulcer, and pain, have revealed the potential for systematic bias in reporting, with under-reporting of some indicators and over-reporting of others.		inconclusive. Caution should be exercised when interpreting the QI results and other sources of evidence of the quality of care processes should be considered in conjunction with QI results.
Johari, K., Kellogg, C., Vazquez, K., Irvine, K., Rahman, A., & Enguidanos, S. (2018). Ratings game: an analysis of nursing home compare and Yelp ratings. BMJ quality & safety, 27(8), 619-624.	To examine Yelp ratings for NHs in California and compares these ratings with NHC ratings.	retrospective secondary data analyses	675 NHs in California between September and November 2016	NHC 5-star ratings vs. ratings from other sources	Correlations between the Yelp and NHC ratings were relatively weak. The Yelp rating was significantly lower than the 5-star NHC rating and the NHC ratings for staffing and QMs. It was significantly higher than the NHC inspection rating.	single state	When consumers rate NHs on Yelp, their ratings differ considerably from NHC ratings
Joyce, N. R., McGuire, T. G., Bartels, S. J., Mitchell, S. L., &	To compare the quality of care following	retrospective secondar	704782 residents with	presence of special care unit	Admission to a facility with an SCU led to a reduction in	instrumental variable approach	Facilities with an SCU provide better quality of

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
Grabowski, D. C. (2018). The impact of dementia special care units on quality of care: An instrumental variables analysis. Health services research, 53(5), 3657-3679.	admission to a NH with and without a dementia special care unit (SCU) for residents with dementia	y data analyses	dementia during 2005–2010	(SCU) for dementia	inappropriate antipsychotics (9.7%), physical restraints (9.6%), pressure ulcers (3.3%), feeding tubes (8.3 %), and hospitalizations (14.7%). No impact was found on the use of indwelling urinary catheters.		care as measured by several validated quality indicators.
Kim, H. M., Banaszak-Holl, J., Kales, H., Mach, J., Blow, F., & McCarthy, J. F. (2013). Trends and predictors of quality of care in VA nursing homes related to serious mental illness. Medical Care, 659-665.	To examine recent trends in quality indicators measuring poor performance of VA NHs and whether the facility-level QIs vary with SMI concentration within the facility	retrospective secondary data analyses	135 Veterans Affairs (VA) NHs between fiscal years 2005-2007	NHs with higher serious mental illness (SMI) concentrations	Higher SMI prevalence was associated with higher odds of physical restraint use and lack of toileting plan. Higher SMI prevalence was also associated with higher frequency of untreated depression	VA NHs	Overall poorer quality was observed at sites with higher SMI concentrations
Kim, S. J., Park, E.-c., Kim, S., Nakagawa, S., Lung, J., Choi, J. B., . . . Kim, K. (2014). The association between quality of care and quality of life in long-	To assess the overall quality of life of long-stay NH residents with preserved cognition, and to examine whether	primary and secondary data analyses	251 long-stay NH residents with preserved cognition in	quality measures and quality of life	Quality of life was associated with physical impairment and depression but not NHC overall star rating and not pain.	resident-level	The 5-star quality rating system did not reflect the quality of life of long-stay NH residents with

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
stay nursing home residents with preserved cognition. Journal of the American Medical Directors Association, 15(3), 220-225.	the NHC 5-star quality rating system reflects the overall quality of life of such residents		32 NHs in Detroit				preserved cognition
Li, Y., Cai, X., & Wang, M. (2019). Social media ratings of nursing homes associated with experience of care and "Nursing Home Compare" quality measures. BMC health services research, 19(1), 260.	To determine if aggregated ratings from 4 popular social media or online review sites (Facebook, Yelp, Google Review, and Caring.com) were associated with family-reported care experience scores, and with CMS' NHC 5-star ratings and other quality measures.	retrospective secondary data analyses	196 NHs in Maryland from July 2015 to July 2017	NHC 5-star ratings vs. ratings from other sources	The overall ratings were 3.11 on average on these sites and 3.03 on the NHC website, with a Pearson correlation of 0.41 ($p < 0.001$) between the 2 sets of ratings. The correlations between the social media rating and survey-based experience-of-care ratings ranged from 0.40 to 0.60, and the correlations between the social media rating and individual NHC quality measures of citations, nurse staffing, and complaints were about 0.35 (in absolute values).	single state	The 5-star ratings collected from 4 social networking sites was correlated with and predictive of the NHC and survey-based experience-of-care measures for Maryland NHs.

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
Mashouri, P., Taati, B., Quirt, H., & Iaboni, A. (2019). Quality Indicators as Predictors of Future Inspection Performance in Ontario Nursing Homes. Journal of the American Medical Directors Association.	To examine whether quality indicators could predict future inspection performance in LTC homes across Ontario, Canada	retrospective secondary data analyses	594 LTC facilities in Ontario, Canada from 2017 to 2018	quality indicators and inspection performance	After running a wide range of models, only a weak relationship was found between quality indicators and future inspection performance. The best-performing model was able to achieve a classification accuracy of 40.1%.Experiencing worsened pain, restraint use, and worsened pressure ulcers were correlated with homes predicted as needing significant improvement. Counterintuitively, improved physical functioning had an inverse relationship with homes predicted as being in good standing.	Canada	Most quality indicators are poor predictors of inspection performance
McGarry, B. E., Joyce, N. R., McGuire, T. G., Mitchell, S. L., Bartels, S. J., & Grabowski, D. C. (2019). Association	To examine the association between the quality of care delivered to NH	retrospective secondar	58571 residents in 12027 NHs	NHs with higher serious mental illness	For individuals with SMI, admission to a high-SMI facility was associated with a 3.7 percentage point	national data	Admission to NHs with high concentrations of residents with SMI is

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
between High Proportions of Seriously Mentally Ill Nursing Home Residents and the Quality of Resident Care. Journal of the American Geriatrics Society, 67(11), 2346-2352.	residents with and without a serious mental illness (SMI) and the proportion of NH residents with SMI.	y data analyses	from 2006-2010	(SMI) concentrations	increase in the probability of feeding tube use relative to individuals admitted to a low-SMI facility. Among individuals without SMI, admission to a high-SMI facility was associated with a 1.7 percentage point increase in the probability of catheter use, a 3.8 percentage point increase in the probability of being hospitalized, and a 2.1 percentage point increase in the probability of having a feeding tube.		associated with worse outcomes for both residents with and without SMI
Mukamel, D. B., Amin, A., Weimer, D. L., Sharit, J., Ladd, H., & Sorkin, D. H. (2016). When patients customize nursing home ratings, choices and rankings differ from the government's	To compare data with 146 residents who used the individualized nursing home compare plus composite measure with the	Demonstration project comparing personalized selection of measure	146 patients and families (42 were patients) who were discharged from hospital to	Difference between measures, weighting, rankings	Almost all users (97%+) selected PT and nurse staffing in their measure; high variability among other measures; <15% chose restraints or catheters; substantial	May not be a feasible approach to VBP	Personalized measures differed enough between individuals and from CMS that such a model should be considered for

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
version. Health Affairs, 35(4), 714-719.	CMS composite measure	s, weighting and subsequent rankings with the 'one size fits all' model	the nursing home		disagreement between CMS and CMSplus		nursing home selection
Nazir, A., Arling, G., Perkins, A. J., & Boustani, M. (2011). Monitoring quality of care for nursing home residents with behavioral and psychological symptoms related to dementia. Journal of the American Medical Directors Association, 12(9), 660-667.	To evaluate the performance of a new QI for the incidence of worsening behaviors in NH residents with behavioral and psychological symptoms association with dementia.	retrospective secondary data analyses	381 Minnesota NHs with 26,165 residents for the first 2 calendar quarters of 2008.	presence of special care unit (SCU) for dementia	The incidence rates of the worsening behavior QI in SCU ranged from 14% in residents with very severe cognitive impairment to 30% in those with moderate cognitive impairment. The incidence QI rates among residents residing in conventional unit ranged from 15% among those with very severe cognitive impairment to 20% among those with moderate cognitive impairment. These differences in QI rates between the 2 units	single state	Resident cognitive function and the facility utility of SCU are associated with worsening behavior QI and should be adjusted for in any NH quality reporting measure.

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
					were statistically significant. After risk adjustment for level of cognitive impairment, number of facilities with SCUs that flagged for problem behaviors dropped from 18.4% to 12.4% and the number of conventional units in the low-risk category from 16.8% to 4.7%.		
Nazir, A., Mueller, C., Perkins, A., & Arling, G. (2012). Falls and nursing home residents with cognitive impairment: new insights into quality measures and interventions. Journal of the American Medical Directors Association, 13(9), 819. e811-819. e816.	To examine the relationship between cognitive impairment, residence on dementia special care units (SCUs) and other resident characteristics and likelihood of residents experiencing new falls in NHs	retrospective secondary data analyses	21,587 residents from 381 Minnesota NHs	presence of special care unit (SCU) for dementia	The likelihood of a new fall had a nonlinear association with cognitive impairment. Compared with residents with normal or mild cognitive impairment, the likelihood of a new fall was significantly higher among residents with moderate cognitive impairment. The risk decreased slightly for residents with more advanced cognitive impairment, whereas the presence of severe	single state	Severity of cognitive impairment and residence on SCU impact fall incidence and should be accounted for in future fall-prevention interventions and quality-reporting indicators and measures.

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
					cognitive impairment was not significantly associated with new falls. Overall the likelihood of new falls was significantly higher for residents on SCUs compared with those on conventional units.		
Netten, A., Trukeschitz, B., Beadle-Brown, J., Forder, J., Towers, A.-M., & Welch, E. (2012). Quality of life outcomes for residents and quality ratings of care homes: is there a relationship? Age and ageing, 41(4), 512-517	To capture social care-related quality of life (SCRQoL) outcomes for residents and investigate the relationship between outcomes and regulator quality ratings of homes	primary data analyses	366 residents of 83 English care homes for older people inspected during 2008	quality measures and quality of life	Care homes were delivering substantial gains in SCRQoL, but were more successful in delivering 'basic' (e.g. personal cleanliness) than higher-order domains (e.g. social participation). Outcomes were associated with quality ratings of residential homes but not of NHs.	England	Future quality indicators need to demonstrate their relationship with quality of life outcomes if they are to be a reliable guide to commissioners and private individuals purchasing care.
Neuman, M. D., Wirtalla, C., & Werner, R. M. (2014). Association between skilled nursing facility quality indicators and hospital readmissions.	To measure the association between SNF performance measures and hospital readmissions	retrospective secondary data analyses	14251 SNFs between September 1, 2009, and August 31, 2010	quality measures and hospitalization	The unadjusted risk of readmission or death was lower at SNFs with better staffing ratings and better facility inspection ratings. Adjustment for resident	national data	Available performance measures were not consistently associated with differences in the adjusted risk

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
JAMA, 312(15), 1542-1551.	among Medicare beneficiaries receiving postacute care at SNFs				factors, facility factors, and the discharging hospital attenuated these associations; small differences were observed in the adjusted risk of readmission or death according to SNF facility inspection ratings.		of readmission or death.
O'Malley, A. J., Caudry, D. J., & Grabowski, D. C. (2011). Predictors of nursing home residents' time to hospitalization. Health services research, 46(1p1), 82-104.	To model the predictors of the time to first acute hospitalization for NH residents, and accounting for previous hospitalizations, model the predictors of time between subsequent hospitalizations.	retrospective secondary data analyses	677 NHs in New York State for the period 1998–2004	quality measures and hospitalization	Pressure ulcers and facility-level deficiencies were associated with a decreased time to first hospitalization, while the presence of advance directives and facility staffing was associated with an increased time. These predictors of the time to first hospitalization model had effects of similar magnitude in predicting the time between subsequent hospitalizations.	single state	Modifiable patient and NH characteristics are associated with the time to first hospitalization and time to subsequent hospitalizations for NH residents

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
Rahman, M., Grabowski, D. C., Intrator, O., Cai, S., & Mor, V. (2013). Serious mental illness and nursing home quality of care. Health services research, 48(4), 1279-1298.	To estimate the effect of a NH's share of residents with a serious mental illness (SMI) on the quality of care	retrospective secondary data analyses	13793 NHs over the period 2000 through 2008	NHs with higher serious mental illness (SMI) concentrations	An increase in the share of SMI NH residents positively affected the hospitalization rate among non-SMI residents and negatively affected staffing skill mix and level. There was no statistically significant effect on inspection-based health deficiencies or the hospitalization rate for SMI residents.	instrumental variables approach	Across the majority of indicators, a greater SMI share resulted in lower NH quality.
Rivera-Hernandez, M., Rahman, M., Mukamel, D. B., Mor, V., & Trivedi, A. N. (2019). Quality of post-acute care in skilled nursing facilities that disproportionately serve black and Hispanic patients. The Journals of Gerontology: Series A, 74(5), 689-697.	To investigate racial and ethnic disparities in the quality of post-acute care in SNFs	retrospective secondary data analyses	649,187 Medicare beneficiaries 65+ from 8,375 SNFs	NHs with higher proportion of African American residents	SNFs with higher fractions of African American patients had worse performance for three publicly reported quality measures: rehospitalization, successful discharge to the community, and the star rating indicator.	national data	Efforts to address disparities should focus attention on NHs that disproportionately serve minority patients and monitor unintended consequences of value-based

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
							payments to SNFs
Saliba, D., Weimer, D. L., Shi, Y., & Mukamel, D. B. (2018). Examination of the new short-stay nursing home quality measures: rehospitalizations, emergency department visits, and successful returns to the community. INQUIRY: The Journal of Health Care Organization, Provision, and Financing, 55	To examine 3 new short-stay quality measures (QMs) — rehospitalizations, emergency department visits, and successful discharges to the community	retrospective secondary data analyses	31, 312 NHs between April 2016 and October 2017	short-stay quality measures	similar to other QMs, performance varies across the country, and that there is very minimal correlation between these 3 new QMs as well as between these QMs and other NHC QMs. Better performance on these QMs tends to be associated with fewer deficiencies, higher staffing and more skilled staffing, nonprofit ownership, and lower proportion of Medicaid residents.	national data	These QMs are important by demonstrating their large variation across the country, suggesting substantial room for improvement
Schapira, M. M., Shea, J. A., Duey, K. A., Kleiman, C., & Werner, R. M. (2016). The nursing home compare report card: perceptions of residents and caregivers regarding	To evaluate the perceived usefulness of the report card to residents and families	Primary data collection, structured interviews	Convenience sample of 35 residents (6) or families (29) newly admitted to the nursing home in the	Perceptions of star ratings, comparisons, and use of the report card for	Positive perception of quality information overall but confusion over how the quality was actually measured and the relationship between domain specific and overall quality score	Convenience sample in a single geographic area	When made aware of the report card people like it, but more clarity is needed for the public to understand the methodology

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
quality ratings and nursing home choice. Health services research, 51, 1212-1228.			Philadelphia area	decision making			
Shwartz, M., Peköz, E. A., Christiansen, C. L., Burgess Jr, J. F., & Berlowitz, D. (2013). Shrinkage estimators for a composite measure of quality conceptualized as a formative construct. Health services research, 48(1), 271-289.	To demonstrate the value of shrinkage estimators when calculating a composite quality measure as the weighted average of a set of individual quality indicators	retrospective secondary data analyses	112 Veterans Health Administration NHs in fiscal years 2005–2008	A composite measure of quality	Usually, shrunken-rate composite scores in 1 year are better able to predict the observed total number of QI events or the observed-rate composite scores in the following year than the initial year observed-rate composite scores.	method to construct a composite measure of quality	Shrinkage estimators can be useful when a composite measure is conceptualized as a formative construct
Snyder, D. J., Kroshus, T. R., Keswani, A., Garden, E. B., Koenig, K. M., Bozic, K. J., . . . Moucha, C. S. (2019). Are Medicare's Nursing Home Compare Ratings Accurate Predictors of 90-Day Complications, Readmission, and Bundle Cost for	to evaluate whether NHC ratings are valid predictors of 90-day complications, readmission, and bundle costs for patients discharged to an SNF after primary total joint	retrospective secondary data analyses	488 patients who discharged to 105 skilled nursing facilities after primary total joint arthroplasty	quality measures and hospitalization	Overall NHC rating was not predictive of 90-day readmission/major complications, >75th percentile postacute cost, or 90-day bundle cost exceeding the target price. SNF health inspection and quality measure ratings were also not predictive of 90-day	patients discharged to an SNF after primary total joint arthroplasty (TJA)	Results of our study suggest that Medicare's NHC tool is not a useful predictor of 90-day costs, complications, or readmissions for SNFs within our health system

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
Patients Undergoing Primary Total Joint Arthroplasty? The Journal of arthroplasty, 34(4), 613-618.	arthroplasty (TJA).				readmission/major complications or bundle performance. A higher SNF staffing rating was independently associated with a decreased odds for >75th percentile 90-day postacute spend and a 90-day bundle cost exceeding the target price but was similarly not predictive of 90-day readmission/ complications.		
Stevenson, D. G., Spittal, M. J., & Studdert, D. M. (2013). Does litigation increase or decrease health care quality? A national study of negligence claims against nursing homes. Medical Care, 51(5), 430.	To assess whether the experience of being sued and incurring litigation costs affects the quality of care subsequently delivered in NHs	retrospective secondary data analyses	1514 NHs between 1998 and 2010	quality measures and negligence litigation	Nearly all combinations of the 3 litigation exposure measures and 9 quality measures (27 models in all) showed an inverse relationship between litigation costs and quality. However only a few of these associations were statistically significant, and the effect sizes were very small.	linked information on negligence claims to indicators of NH quality drawn from two national data sets	Tort litigation does not increase the NH quality performance, and may decrease it slightly.

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
Studdert, D. M., Spittal, M. J., Mello, M. M., O'Malley, A. J., & Stevenson, D. G. (2011). Relationship between quality of care and negligence litigation in nursing homes. New England Journal of Medicine, 364(13), 1243-1250	To investigate whether high-quality NHs are less likely to be sued for negligence than their low-performing counterparts	retrospective secondary data analyses	1465 NHs between 1998 and 2006	quality measures and negligence litigation	NHs with more deficiencies and those with more serious deficiencies had higher odds of being sued; this was also true for NHs that had more residents with weight loss and with pressure ulcers. The odds of being sued were lower in NHs with more nurse's aide–hours per resident-day. However, all these effects were relatively small.	linked information on tort claims to 10 indicators of NH quality drawn from two national data sets	The best-performing NHs are sued only marginally less than the worst-performing ones. Such weak discrimination may subvert the capacity of litigation to provide incentives to deliver safer care.
Troyer, J. L., & Sause, W. (2013). Association between traditional nursing home quality measures and two sources of nursing home complaints. Health services research, 48(4), 1256.	To test for an association between traditional NH quality measures and two sources of resident- and caregiver-derived NH complaints	retrospective secondary data analyses	379 NHs in North Carolina from 2002-2006	quality measures and NH complaints	There is little relationship between MDS-QIs and complaints. Ombudsman complaints and inspection violations are generally unrelated, but there is a positive relationship between state certification agency complaints and inspection violations.	single state	Ombudsman and state certification agency complaint data are resident and caregiver-derived quality measures that are distinctive from and complement traditional

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
							quality measures.
Werner, R. M., Konetzka, R. T., & Kim, M. M. (2013). Quality improvement under nursing home compare: the association between changes in process and outcome measures. Medical Care, 51(7), 582.	To test the extent to which improvements in outcomes of care are explained by changes in NH processes	retrospective secondary data analyses	16,623 NHs from 2000 to 2009	process and outcome measures	Of the 5 outcome measures examined, only improvements in the percentage of NH residents in moderate or severe pain were associated with changes in NH processes of care. Furthermore, these changes in the measured process of care explained only a small part of the overall improvement in pain prevalence.	national data	A large portion of the improvements in NH outcomes were not associated with changes in measured processes of care suggesting that processes of care typically measured in NHs do little to improve NH performance on outcome measures.
Williams, A., Straker, J. K., & Applebaum, R. (2016). The nursing home five star rating: How does it compare to resident and family views of care? The Gerontologist, 56(2), 234-242.	To compares the CMS star rating system to NH satisfaction data reported by residents and their families in Ohio	retrospective secondary data analyses	918 NHs in Ohio	quality measures and satisfaction	Many 5-star NHs had moderate to very low consumer satisfaction (54% compared with resident satisfaction and 41% compared with family satisfaction), and many 1-star NHs had high to very high	single state	Findings indicate that the star rating system does not adequately reflect consumer satisfaction

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
					consumer satisfaction (19% compared with resident satisfaction and 20.0% compared with family satisfaction)		
Xu, D., Kane, R., & Arling, G. (2019). Relationship between nursing home quality indicators and potentially preventable hospitalisation. BMJ quality & safety, 28(7), 524-533.	To examine the relationship between quality indicators and overall and potentially preventable hospitalizations among Medicaid beneficiaries aged 65 years and older receiving care at NHs in Minnesota.	retrospective secondary data analyses	20 518 Medicaid beneficiaries aged 65+ in 345 NHs during the 2011-2012 period	quality measures and hospitalization	Available quality indicators were not strongly or consistently associated with the risk of hospitalization. Among these 23 quality indicators, five quality indicators (antipsychotics without a diagnosis of psychosis, unexplained weight loss, pressures sores, bladder continence and activities of daily living [ADL] dependence) were related significantly to hospitalization and only four quality indicators (antipsychotics without a diagnosis of psychosis, unexplained weight loss, ADL dependence and urinary tract infections) were related	single state	Although general quality indicators can be informative about overall NH performance, only selected quality indicators appear to tap dimensions of clinical quality directly related to hospitalizations.

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
					to potentially preventable hospitalization.		
Yuan, Y., Lapane, K. L., Baek, J., Jesdale, B. M., & Ulbricht, C. M. (2019). Nursing Home Star Ratings and New Onset of Depression in Long-Stay Nursing Home Residents. Journal of the American Medical Directors Association, 20(10), 1335-1339. e1310.	To examine the association between NH quality and new onset of depression and severity of depressive symptoms in a national cohort of long-stay NH residents	retrospective secondary data analyses	29,837 long-stay residents without indicators of depression in 13,921 NHs	quality measures and depression	Using minimal depressive symptoms as the reference, residents in NHs with 5-star overall ratings were 12% less likely than those in 3-star NHs to experience mild and 31% less likely to experience moderate symptoms. In NHs with 1-star staffing compared to 3-star, residents had 37% higher odds of moderate symptoms and 57% higher odds of moderately severe to severe depressive symptoms. The odds of any above-minimal depressive symptoms decreased as quality measure ratings increased.	national data	Lower NH quality ratings were associated with more severe depressive symptoms.

Chapter 2

Analysis of Minnesota's Nursing Facility Clinical Quality Indictors

Dongjuan Xu, PhD MSN RN

Executive Summary

The Minnesota Nursing Home Report Card provides two clinical quality indictor (QI) ratings: one focused on the quality of care during long-term stays (LS) with 19 indicators, and one focused on the quality of care during short-term stays (SS) with 2 indicators. Currently, face validity and expert opinions are employed to group the 19 long-stay QIs into 10 different domains or aspects of care. However, we do not know whether these domains are supported by the data. Under the current scoring program, the best performing 20% of facilities statewide get full points on each QI, the worst performing 10% get no points, and the rest are sorted and given a prorated point value. However, some QIs may not discriminate very well between facilities. The main objective was to explore the possibility of reducing the number of clinical QIs using exploratory factor analysis (EFA), Cronbach's alpha, correlation, scatter plots, descriptive and trends analysis. Risk-adjusted facility-level QIs including 19 long-stay QIs and 2 short-stay QIs over the 2012-2019 period (four quarters in each year) were used. The number of nursing facilities in each quarter ranged from 369 to 382.

The EFA results suggest that it is reasonable to posit 5 underlying dimensions or domains of the 19 long-stay facility-level QIs: incontinence (2 QIs), no toileting plan for incontinence (2 QIs), physical functioning (5 QIs), restraints and behavioral symptoms (5 QIs), and care for specific conditions (5 QIs). The five factors explained 44.03% of variance in QIs. The Cronbach's alpha for these 5 factors were 0.79, 0.81, 0.47, 0.41, and 0.36, which was acceptable. Characterizing facility performance in parsimonious but meaningful ways based on the dimensionality of QIs would simplify interpretation. Moreover, because of the moderate correlations between QIs, it is reasonable to combine the two QIs "incidence of worsening or serious bladder incontinence" and "incidence of worsening or serious bowel incontinence" (correlation coefficient: 0.657) into one QI, combine the two QIs "prevalence of occasional to full bladder incontinence without a toileting plan" and "prevalence of occasional to full bowel incontinence without a toileting plan" (correlation coefficient: 0.683) into one QI, and combine the two QIs "incidence of worsening or serious functional dependence" and "incidence of worsening or serious mobility dependence" (correlation coefficient: 0.508) into one QI.

We also found that 8 QIs with too little variance and floor effects (a large concentration of very low scores on the measurement scale), which may not discriminate very well between facilities under the current scoring program. There were additional 4 QIs with either floor or ceiling effects, although they had relatively large variance. For example, with the ceiling effect, facilities with the worst performing were so close to the median facilities, they were not very distinguishable. When taking measurement error into account, there was almost no difference in the performance. The point threshold in the current scoring program should be re-adjusted, so that the points are a better reflection of facility performance. Moreover, since the thresholds were based on percentiles or distribution, they would move with overall QI trends. A facility may be improving in its QI rate, but since others were improving as well, that facility would not get any higher points. It would be interesting to further analyze whether the thresholds should be fixed, that is based on the same QI rates over time.

1. Background

The Report Card provides two clinical quality indictor (QI) ratings: one focused on the quality of care during long-term stays (LS) with 19 indicators, and one focused on the quality of care during short-term stays (SS) with 2 indicators. These QIs are risk adjusted to account for differences between the types of residents served in nursing homes (NHs). Examples of the adjustors used are, but are not limited to: age, gender, cognitive performance (mental functioning), Alzheimer's disease, stroke, and ADL ability (Minnesota Nursing Home Report Card Technical User Guide).

2. Objective

The main objective was to explore the possibility of reducing the number of clinical quality indicators (QIs). We are working with the current QIs as defined. We are not evaluating the need for new QIs or major re-defining of the current QIs.

3. Data and Methods

Risk-adjusted facility-level QIs including 19 long-stay QIs and 2 short-stay QIs over the 2012-2019 period (four quarters in each year) were used. The number of NHs in each quarter ranged from 369 to 382. Exploratory factor analysis (EFA), Cronbach's alpha, correlation, scatter plots, descriptive and trends analysis were conducted for this report.

4. Results

4.1 Dimensionality or Domains of NH Quality Indicators

Table 1 shows the 19 long-stay QIs in the Report Card. One QI "incidence of walking as well or better than previous assessment" is a positively framed measure while the rest QIs are negative measures. Performance on QIs captures the differences in certain aspects of quality. Currently, face validity and expert opinions are employed to group the 19 long-stay QIs into 10 different domains or aspects of care. However, we do not know whether these domains are supported by the data. Exploratory factor analysis is a statistical method used to identify a set of latent constructs underlying a battery of measured variables. In this report, EFA was used to investigate the dimensionality of NH quality which may usefully summarize the multiplicity of QIs. Principal component factor methods with orthogonal rotation was used in the EFA. In addition, Cronbach's alpha was calculated to assess the internal consistency of QIs loaded on the same factor or domain. Internal consistency, one of the measure to assess reliability, measures whether several QIs that propose to measure the same general quality construct produce similar scores.

Table 1. Domains of long-stay quality indicators in Report Card

Domain	19 Long-stay Quality Indicators	Full Points
Psychosocial	Incidence of Worsening or Serious Resident Behavior Problems	5
	Prevalence of Depressive Symptoms	5
Quality of Life	Prevalence of Physical Restraints	10
Continence	Incidence of Worsening or Serious Bowel Incontinence	2
	Incidence of Worsening or Serious Bladder Incontinence	2
	Prevalence of Occasional to Full Bladder Incontinence Without a Toileting Plan	2

Domain	19 Long-stay Quality Indicators	Full Points
Infections	Prevalence of Occasional to Full Bowel Incontinence Without a Toileting Plan	2
	Prevalence of Indwelling Catheters	2
	Prevalence of Urinary Tract Infections	5
	Prevalence of Infections	5
Accidents	Prevalence of Falls with Major Injury	10
Nutrition	Prevalence of Unexplained Weight Loss	10
Skin Care	Prevalence of Pressure Sores in High-Risk Residents	10
Psychotropic Drugs	Prevalence of Antipsychotics Without a Diagnosis of Psychosis	10
Physical Functioning	Incidence of Walking as Well or Better than Previous Assessment	2.5
	Incidence of Worsening or Serious Functional Dependence	2.5
	Incidence of Worsening or Serious Mobility Dependence	2.5
	Incidence of Worsening or Serious Range of Motion Limitation	2.5
Pain	Prevalence of Residents who Report Moderate to Severe Pain	10

Table 2 presents the EFA results of 19 long-stay QIs. There were five factors or domains underlying the 19 QIs:

- **Factor/Domain 1: incontinence, including 2 QIs:**
  - **Incidence of worsening or serious bowel incontinence**
  - **Incidence of worsening or serious bladder incontinence**
- **Factor/Domain 2: no toileting plan for incontinence, including 2 QIs:**
  - **Prevalence of occasional to full bowel incontinence without a toileting plan**
  - **Prevalence of occasional to full bladder incontinence without a toileting plan**
- **Factor/Domain 3: physical functioning, including 5 QIs:**
  - **Incidence of walking as well or better than previous assessment**
  - **Incidence of worsening or serious functional dependence**
  - **Incidence of worsening or serious mobility dependence**
  - **Incidence of worsening or serious range of motion limitation**
  - **Prevalence of falls with major injury**
- **Factor/Domain 4: restraints and behavioral symptoms, including 5 QIs:**
  - **Incidence of worsening or serious resident behavior problems**
  - **Prevalence of depressive symptoms**
  - **Prevalence of physical restraints**
  - **Prevalence of antipsychotic medications without a diagnosis of psychosis**

- **Prevalence of infections**

- **Factor/Domain 5: care for specific conditions, including 5 QIs:**
  - **Prevalence of moderate to severe pain**
  - **Prevalence of pressure sores in high risk residents**
  - **Prevalence of unexplained weight loss**
  - **Prevalence of indwelling catheters**
  - **Prevalence of urinary tract infections**

The five factors explained 44.03% of variance in QIs. The eigenvalues for these 5 factors were 2.48, 2.08 1.48, 1.23, and 1.10. The signs of factor loading were consistent and as expected in each factor, with the QI "incidence of walking as well or better than previous assessment" has a negative loading score. Contrary to expectation, the QI "prevalence of infections" loaded on factor 4 "restraints and behavioral symptoms" and did not load on the same factor with the QI "prevalence of urinary tract infections".

The Cronbach's alpha for these 5 factors were 0.79, 0.81, 0.47, 0.41, and 0.36. Usually, the internal consistency is acceptable when Cronbach's alpha is 0.7 or greater, although some researchers suggested 0.4 is acceptable for the reliability of MDS items (Hawes et al, 1995).

In summary, the EFA results suggest that it is reasonable to posit 5 underlying dimensions or domains of the 19 long-stay facility-level QIs. Characterizing facility performance in parsimonious but meaningful ways based on the dimensionality of QIs would simplify interpretation. It offers an opportunity to create summary factors or domain scores of quality performance in certain aspects of care.

Table 2. EFA results of 19 long-stay QIs

19 Long-Stay Quality Indicators	Factor1	Factor2	Factor3	Factor4	Factor5
CNTA: Adjusted I of Worsening or Serious Bowel Incontinence (LS)	0.818	0.126	0.124	0.022	-0.017
CNTB: Adjusted I of Worsening or Serious Bladder Incontinence (LS)	0.816	0.166	0.066	0.003	0.019
CNTF: Adjusted P of Occasional to Full Bowel Incontinence w/o a Toileting Plan (LS)	0.071	0.867	-0.054	0.002	-0.003
CNTE: Adjusted P of Occasional to Full Bladder Incontinence w/o a Toileting Plan (LS)	0.181	0.878	0.011	0.013	-0.036
WALX: Adjusted I of Walking as Well or Better than on Previous Assessment (LS)	-0.424	-0.113	-0.305	0.331	-0.083
ADLA: Adjusted I of Worsening or Serious Functional Dependence (LS)	0.312	-0.048	0.713	0.080	0.023
MOBA: Adjusted I of Worsening or Serious Mobility Dependence (LS)	0.089	-0.046	0.685	0.111	0.039
ROMA: Adjusted I of Worsening or Serious Range of Motion Limitation (LS)	-0.079	0.053	0.351	0.056	0.170
FAL1: Adjusted P of Falls with Injury (LS)	-0.227	0.042	0.409	-0.175	-0.023
BEHA: Adjusted I of Worsening or Serious Resident Behavior Problems (LS)	-0.099	0.069	0.209	0.598	-0.266
MOD1: Adjusted P of Depressive Symptoms (LS)	-0.074	0.052	0.117	0.538	0.080
RES1: Adjusted P of Physical Restraints (LS)	0.044	-0.154	-0.114	0.401	0.253
DRG1: Adjusted P of Antipsychotic Medications Without a Diagnosis of Psychosis (LS)	0.168	0.039	0.016	0.596	0.123
INFX: Adjusted P of Infections (LS)	0.062	-0.180	0.263	0.310	0.083
PAI3: Adjusted P of Moderate to Severe Pain (LS)	-0.200	0.312	0.224	0.196	0.302
PRUB: Adjusted P of Pressure Sores in High Risk Residents (LS)	0.202	0.131	0.026	-0.044	0.521
WGT1: Adjusted P of Unexplained Weight Loss (LS)	0.149	-0.019	0.129	0.247	0.361
CAT2: Adjusted P of Indwelling Catheters (LS)	0.018	-0.050	0.016	-0.104	0.679

19 Long-Stay Quality Indicators	Factor1	Factor2	Factor3	Factor4	Factor5
CNT4: Adjusted P of Urinary Tract Infections (LS)	-0.265	-0.185	0.089	0.238	0.524

Besides the 19 long-stay QIs, there are 2 short-stay QIs in the Report Card: prevalence of moderate to severe pain and prevalence of new or worsening pressure sores. Table 3 presents the EFA results of 21 QIs. There were 6 factors or domains underlying the 21 QIs. As expected, the short- and long-stay pain QIs loaded together on the same factor (Factor 3) and the short- and long-stay pressure sore QIs loaded together on the same factor (Factor 6). The six factors explained 47.46% of variance in QIs. The eigenvalues for these 6 factors were 2.49, 2.28, 1.65, 1.27, 1.20 and 1.07. The Cronbach's alpha for these 6 factors were 0.79, 0.81, 0.75, 0.49, 0.41, and 0.36.

Table 3. EFA results of 21 QIs

21 Quality Indicators	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
CNTA: Adjusted I of Worsening or Serious Bowel Incontinence (LS)	0.847	0.101	-0.021	0.104	0.011	-0.029
CNTB: Adjusted I of Worsening or Serious Bladder Incontinence (LS)	0.849	0.146	-0.009	0.041	-0.007	0.033
CNTE: Adjusted P of Occasional to Full Bladder Incontinence w/o a Toileting Plan (LS)	0.187	0.873	0.092	0.025	0.009	-0.035
CNTF: Adjusted P of Occasional to Full Bowel Incontinence w/o a Toileting Plan (LS)	0.053	0.881	0.023	-0.019	0.019	0.014
PAI3: Adjusted P of Moderate to Severe Pain (LS)	0.004	0.079	0.866	0.028	0.016	0.028
PAI2: Adjusted P of Moderate to Severe Pain (SS)	-0.034	0.035	0.857	0.009	0.054	0.044
WALX: Adjusted I of Walking as Well or Better than on Previous Assessment (LS)	-0.391	-0.116	0.026	-0.351	0.359	-0.032
ADLA: Adjusted I of Worsening or Serious Functional Dependence (LS)	0.261	-0.031	0.028	0.754	0.070	0.010
MOBA: Adjusted I of Worsening or Serious Mobility Dependence (LS)	-0.027	0.022	0.017	0.788	0.090	0.077
FAL1: Adjusted P of Falls with Injury (LS)	-0.093	-0.082	0.160	0.264	-0.126	-0.155
ROMA: Adjusted I of Worsening or Serious Range of Motion Limitation (LS)	0.049	-0.067	0.186	0.227	0.059	0.065
BEHA: Adjusted I of Worsening or Serious Resident Behavior Problems (LS)	-0.076	0.067	0.094	0.165	0.610	-0.267
MOD1: Adjusted P of Depressive Symptoms (LS)	0.054	-0.048	0.214	0.008	0.522	-0.008
RES1: Adjusted P of Physical Restraints (LS)	-0.065	-0.082	0.008	0.024	0.342	0.298
DRG1: Adjusted P of Antipsychotic Medications Without a Diagnosis of Psychosis (LS)	0.025	0.145	-0.067	0.178	0.569	0.220

21 Quality Indicators	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
INFX: Adjusted P of Infections (LS)	0.219	-0.299	0.133	0.098	0.360	0.008
PRUA: Adjusted P of New or Worsening Pressure Sores (SS)	0.007	-0.024	0.230	-0.099	0.180	0.397
PRUB: Adjusted P of Pressure Sores in High Risk Residents (LS)	0.232	0.100	0.056	0.009	-0.043	0.540
WGT1: Adjusted P of Unexplained Weight Loss (LS)	0.151	-0.033	0.068	0.158	0.247	0.311
CAT2: Adjusted P of Indwelling Catheters (LS)	-0.054	-0.026	0.041	0.130	-0.146	0.641
CNT4: Adjusted P of Urinary Tract Infections (LS)	-0.252	-0.227	0.167	0.098	0.225	0.456

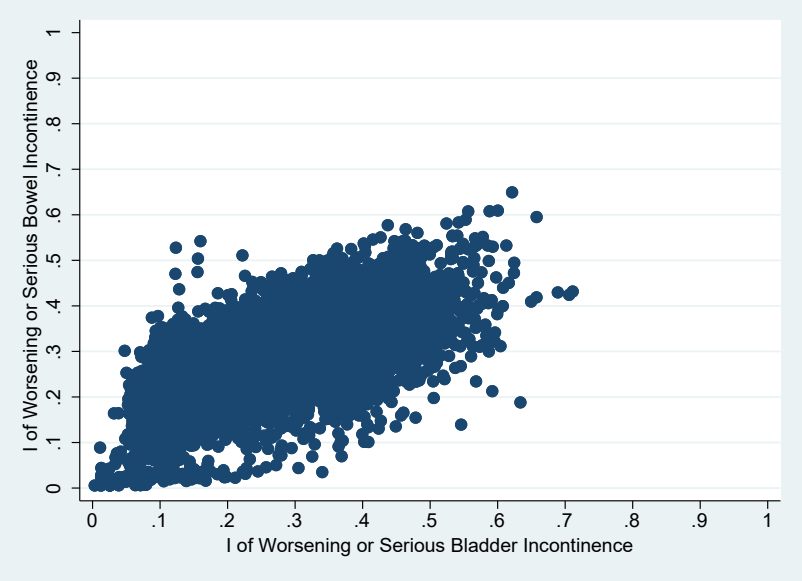
4.2 Further Combining QIs

If a QI is highly correlated with another QI, it may be reasonable to combine them into one QI. Pearson correlation coefficients and scatter plots were used for analysis. As shown in Table 4, The two long-stay QIs "incidence of worsening or serious bladder incontinence" and "incidence of worsening or serious bowel incontinence" has a correlation coefficient of 0.657 (0.7 or above indicating highly correlated). As presented in the scatter plot (Figure 1), if a facility had a high incidence of bowel incontinence, the facility would have a corresponding high incidence of bladder incontinence. Moreover, the two QIs loaded on the same factor (Table 2 and Table 3). It may be reasonable to combine these two QIs into one QI: "incidence of worsening or serious bowel or bladder incontinence".

Table 4. Correlation coefficients between specific QIs related to incontinence

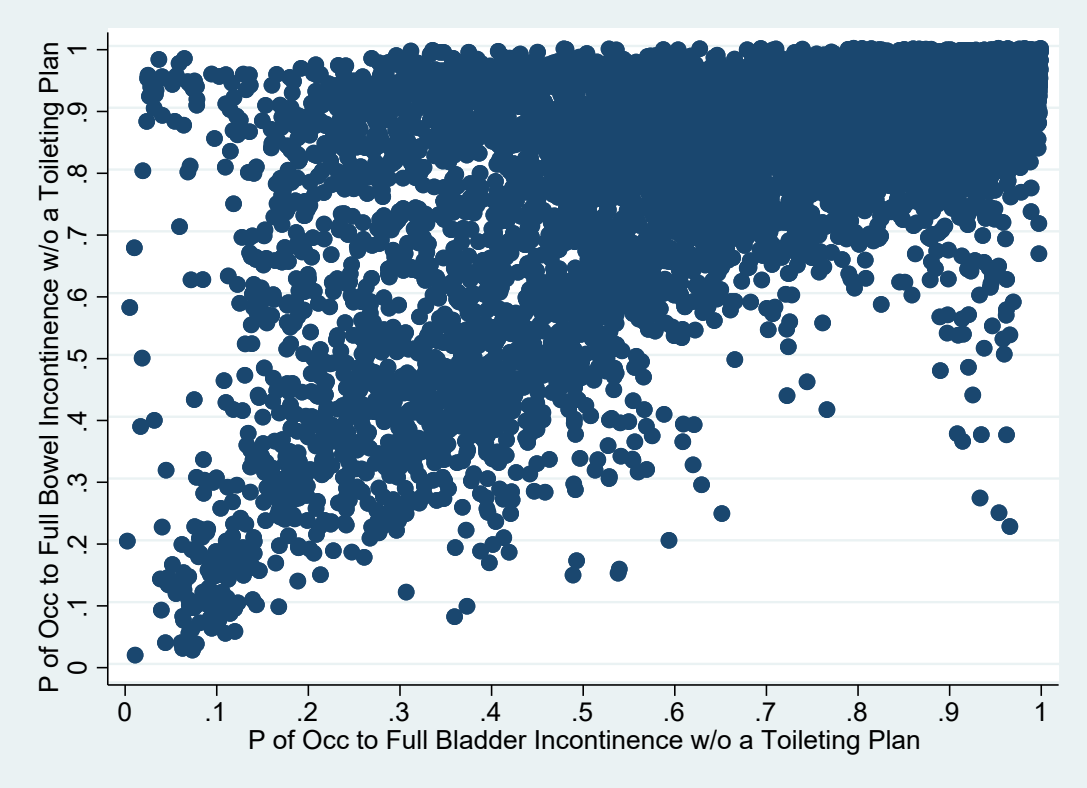
	CNTA	CNTB	CNTE	CNTF
CNTA: Adjusted I of Worsening or Serious Bowel Incontinence (LS)	1			
CNTB: Adjusted I of Worsening or Serious Bladder Incontinence (LS)	0.657	1		
CNTE: Adjusted P of Occasional to Full Bladder Incontinence w/o a Toileting Plan (LS)	0.222	0.276	1	
CNTF: Adjusted P of Occasional to Full Bowel Incontinence w/o a Toileting Plan (LS)	0.157	0.173	0.683	1

Figure 1. Scatter plot of two QIs: bowel incontinence and bladder incontinence



The two long-stay QIs "prevalence of occasional to full bladder incontinence without a toileting plan" and "prevalence of occasional to full bowel incontinence without a toileting plan" has a correlation coefficient of 0.683 (Table 4). These two QIs loaded on the same factor (Table 2 and Table 3). Although the scatter plot does not show a linear relationship (Figure 2), it might be reasonable to combine these two QIs into one QI: "prevalence of occasional to full bladder or bowel incontinence without a toileting plan".

Figure 2. Scatter plot of two QIs: bowel and bladder incontinence without a toileting plan

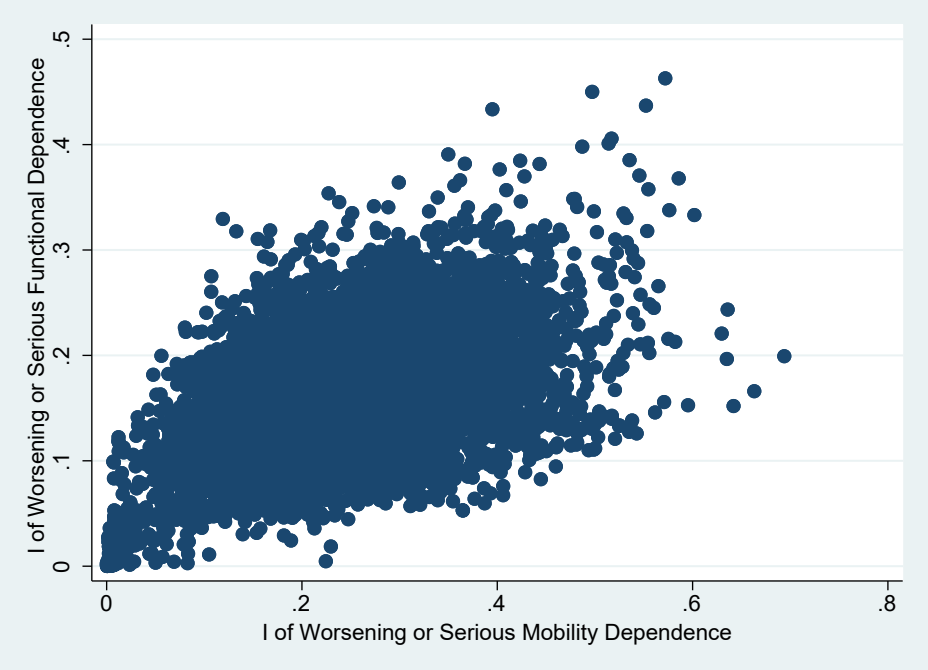


Two long-stay physical functioning QIs "incidence of worsening or serious functional dependence" and "incidence of worsening or serious mobility dependence" are moderately correlated (correlation coefficient 0.508, Table 5 and scatter plot in Figure 3). These two QIs loaded on the same factor (Table 2 and Table 3). It may be reasonable to combine these two QIs into one QI: "incidence of worsening or serious functional or mobility dependence".

Table 5. Correlation coefficients between specific QIs related to physical functioning

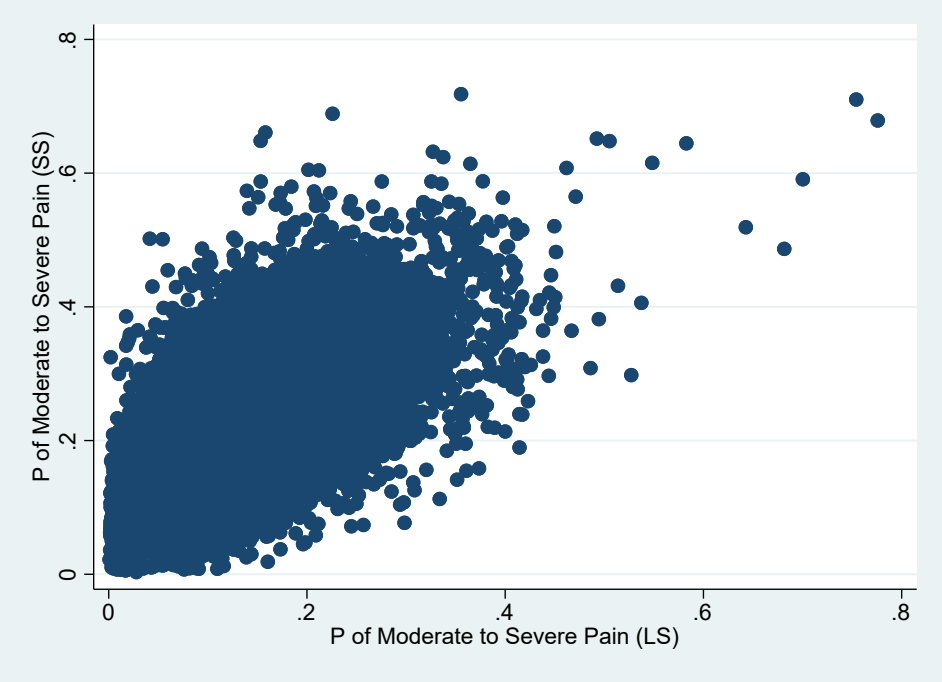
	WALX	ADLA	MOBA	ROMA	FAL1
WALX: Adjusted I of Walking as Well or Better than on Previous Assessment (LS)	1				
ADLA: Adjusted I of Worsening or Serious Functional Dependence (LS)	-0.303	1			
MOBA: Adjusted I of Worsening or Serious Mobility Dependence (LS)	-0.170	0.508	1		
ROMA: Adjusted I of Worsening or Serious Range of Motion Limitation (LS)	-0.082	0.152	0.139	1	
FAL1: Adjusted P of Falls with Injury (LS)	-0.013	0.045	0.045	0.012	1

Figure 3. Scatter plot of two QIs: functional dependence and mobility dependence



Although the short- and long-stay pain QIs are moderately correlated with correlation coefficient 0.600 (scatter plot in Figure 4) and these two QIs loaded on the same factor (Table 3), they should keep as separate QI because the QIs assessed two different population: short-stay residents and long-stay residents.

Figure 4. Scatter plot of two short- and long-stay pain QIs



4.3 Potential Individual QIs That May Not Discriminate Very Well Between Facilities

One of the quality measure evaluation criteria recommended by the National Quality Forum is substantial potential for improvement. There is considerable variation or overall less-than-optimal performance in the quality of care across health care providers. If there is too little variation, the

measure cannot discriminate facilities very well. Descriptive analysis was conducted to summarize the QIs including mean, standard deviation, minimum value, maximum value, range and variance. As shown in Table 6, there are 8 QIs with variance less than 0.001.

- **Prevalence of physical restraints (long-stay)**
- **Prevalence of new or worsening pressure sores (short-stay)**
- **Prevalence of pressure sores in high risk residents (long-stay)**
- **Prevalence of indwelling catheters (long-stay)**
- **Prevalence of infections (long-stay)**
- **Prevalence of falls with injury (long-stay)**
- **Prevalence of urinary tract infection (long-stay)**
- **Prevalence of unexplained weight loss (long-stay)**

The scoring program used for the Nursing Home Report Card compares facilities against each other and combine QIs with very different ranges into one score. Basically, the best performing 20% of facilities statewide get full points on each QI, the worst performing 10% get no points, and the rest are sorted and given a prorated point value used the following equation. As shown in Table 1, there are 10 domains of 19 long-stay QIs. Each long-stay domain is assigned 10 points and within each domain the points are distributed equally (Table 1). Finally, the domains are added into a total score for a maximum 100 points.

$$\frac{(\text{Adjusted Facility Rate} - \text{Rate for No Points})}{(\text{Rate for Full Points} - \text{Rate for No Points})} \text{ x Possible QI Points}$$

The scoring program is kind of normalizing the QI rate (prevalence or incidence). If the QI rate has an approximate normal distribution, the scoring program has no concerns. However, if the distribution of QI rate is highly skewed with a floor effect (a large concentration of very low scores on the measurement scale) or ceiling effect (a large concentration of very high scores on the measurement scale), the scoring program will distort and exaggerate the differences in the low scores or high scores.

First, the distribution of each QI was explored, starting with the 8 QIs with too little variance, then the QIs with skewed distributions (floor or ceiling effects), and the QIs with approximate normal distributions.

Second, the trends (line graphs) of each QI rate over the 2012-2019 period were drawn and compared among the average score of the worst performing 10% of facilities, the median, and the average score of the best performing 20% of facilities. The line graphs indicate thresholds for achieving 0, 5, or 10 points (QIs: "prevalence of physical restraints", "prevalence of pressure sores in high risk residents", "prevalence of falls with injury", and "prevalence of unexplained weight loss"), thresholds for achieving 0, 2.5, or 5 points (QIs: "prevalence of new or worsening pressure sores", "prevalence of infections", and "prevalence of urinary tract infection"), or thresholds for achieving 0, 1, or 2 points (QI: "prevalence of indwelling catheters"). The spread between the lines gives an indication of how well the points are distributed under the current scoring program. For example, regarding the QIs with the line graphs indicating thresholds for achieving 0, 5, or 10 points, facilities getting 5 points (median) are so close to the 10-point threshold (the best performing 20% of facilities), they are hardly distinguishable. When taking measurement error into account, there is hardly any difference in the performance of the 10 and 5-point facilities on those QIs. Thus, the QIs may not discriminate very well between facilities. The point threshold should be re-adjusted, so that the points are a better reflection of facility performance.

Table 6. The descriptive results of 21 QIs (ranked by SD from largest to smallest)

21 Long- and Short-Stay Quality Indicators	N	Mean	SD	Min	Max	Range	Variance
CNTE: Adjusted P of Occasional to Full Bladder Incontinence w/o a Toileting Plan (LS)	11,584	0.749	0.233	0.002	0.999	0.997	0.0544
CNTF: Adjusted P of Occasional to Full Bowel Incontinence w/o a Toileting Plan (LS)	11,432	0.848	0.168	0.020	0.998	0.978	0.0282
WALX: Adjusted I of Walking as Well or Better than on Previous Assessment (LS)	11,688	0.691	0.115	0.136	1	0.863	0.0133
CNTB: Adjusted I of Worsening or Serious Bladder Incontinence (LS)	11,717	0.271	0.103	0.003	0.711	0.708	0.0107
PAI2: Adjusted P of Moderate to Severe Pain (SS)	11,446	0.249	0.103	0.003	0.718	0.715	0.0105
MOBA: Adjusted I of Worsening or Serious Mobility Dependence (LS)	11,726	0.238	0.090	<0.001	0.693	0.693	0.0080
CNTA: Adjusted I of Worsening or Serious Bowel Incontinence (LS)	11,726	0.283	0.081	0.005	0.649	0.644	0.0066
PAI3: Adjusted P of Moderate to Severe Pain (LS)	11,725	0.153	0.081	0.001	0.776	0.775	0.0066
ROMA: Adjusted I of Worsening or Serious Range of Motion Limitation (LS)	11,690	0.115	0.077	<0.001	0.618	0.617	0.0059
BEHA: Adjusted I of Worsening or Serious Resident Behavior Problems (LS)	11,729	0.125	0.067	0.001	0.622	0.621	0.0045
DRG1: Adjusted P of Antipsychotic Medications Without a Diagnosis of Psychosis (LS)	11,558	0.082	0.063	0.001	0.702	0.701	0.0039
ADLA: Adjusted I of Worsening or Serious Functional Dependence (LS)	11,726	0.157	0.053	<0.001	0.463	0.463	0.0028
MOD1: Adjusted P of Depressive Symptoms (LS)	11,728	0.051	0.047	<0.001	0.688	0.688	0.0022
WGT1: Adjusted P of Unexplained Weight Loss (LS)	11,727	0.052	0.030	0.001	0.258	0.258	0.0009
CNT4: Adjusted P of Urinary Tract Infections (LS)	11,727	0.041	0.030	0.001	0.268	0.268	0.0009
FAL1: Adjusted P of Falls with Injury (LS)	10,931	0.034	0.027	0.001	0.239	0.238	0.0007
INFX: Adjusted P of Infections (LS)	11,727	0.028	0.027	0.001	0.373	0.373	0.0007
CAT2: Adjusted P of Indwelling Catheters (LS)	11,722	0.027	0.024	<0.001	0.185	0.185	0.0006

21 Long- and Short-Stay Quality Indicators	N	Mean	SD	Min	Max	Range	Variance
PRUB: Adjusted P of Pressure Sores in High Risk Residents (LS)	11,520	0.039	0.020	0.001	0.165	0.164	0.0004
PRUA: Adjusted P of New or Worsening Pressure Sores (SS)	11,677	0.020	0.018	<0.001	0.216	0.216	0.0003
RES1: Adjusted P of Physical Restraints (LS)	11,732	0.006	0.013	<0.001	0.200	0.199	0.0002

4.3.1 Prevalence of Physical Restraints (Long-Stay)

As presented in Figure 5, the distribution of the QI "prevalence of physical restraints" is highly skewed with a floor effect (mean: 0.006; variance: 0.0002). Figure 6 shows that there is almost no difference between the median and the average score of the worst performing 10% of facilities. However, using the current scoring program, the worst performing 10% of facilities got 0 point, the median got about 5 points, and the best performing 20% of facilities got 10 points. Obviously the scoring program distorted the actual differences of QI rate and forcedly assigned an exaggerated point differences.

This QI defines the percent of long-stay residents who were physically restrained. Because of too little variance, on average the difference between the best performing and the worst performing facilities in the number of residents with physical restraints is less than 6 per 100 long-stay residents (Figure 6). Starting at 2017, as the worst performing10% of facilities made improvements in this QI, the difference was less than 2 per 100 long-stay residents. Does this number difference have practical significance? It may be better to set a threshold percent for physical restraint. If a facility is above the threshold percent, it will get a flag on this QI.

Figure 5. The distribution of QI: physical restraints

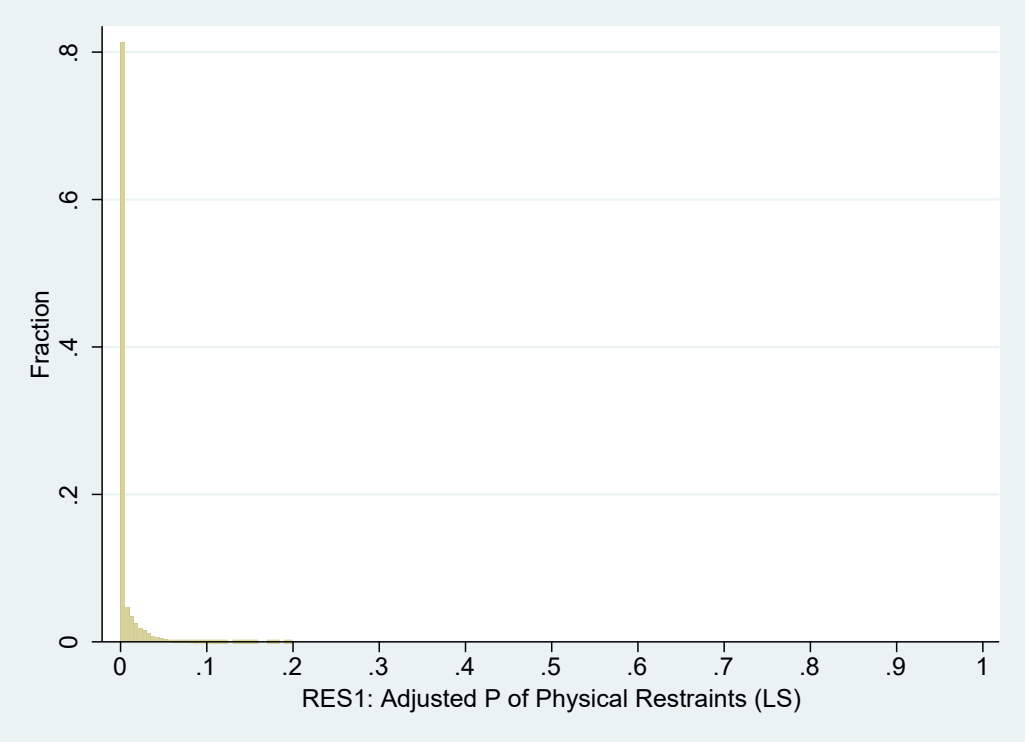
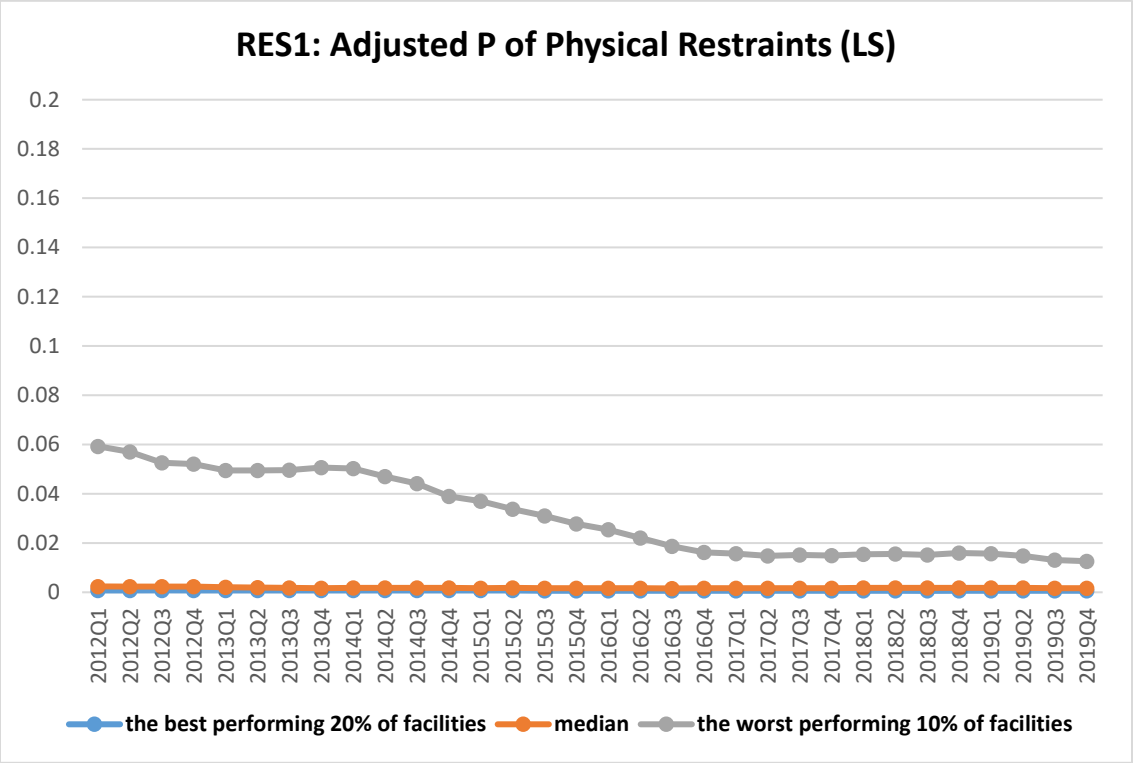


Figure 6: The trends of QI: physical restraints



4.3.2 Prevalence of New or Worsening Pressure Sores (Short-Stay)

As presented in Figure 7, the distribution of the QI "prevalence of new or worsening pressure sores" is highly skewed with a floor effect (mean: 0.020; variance: 0.0003). This is the percent of short-stay residents who have developed pressure sores or who had pressure sores that got worse since admission. Because of too little variance, on average the difference between the best performing 20% of facilities and the worst performing 10% of facilities in the number of residents with new or worsening pressure sores is less than 7 per 100 short-stay residents (Figure 8). Does this number difference have practical significance? It may be better to set a threshold percent for new or worsening pressure sore among short-stay residents. If a facility is above the threshold percent, it will get a flag on this QI.

Figure 7. The distribution of QI: new or worsening pressure sores

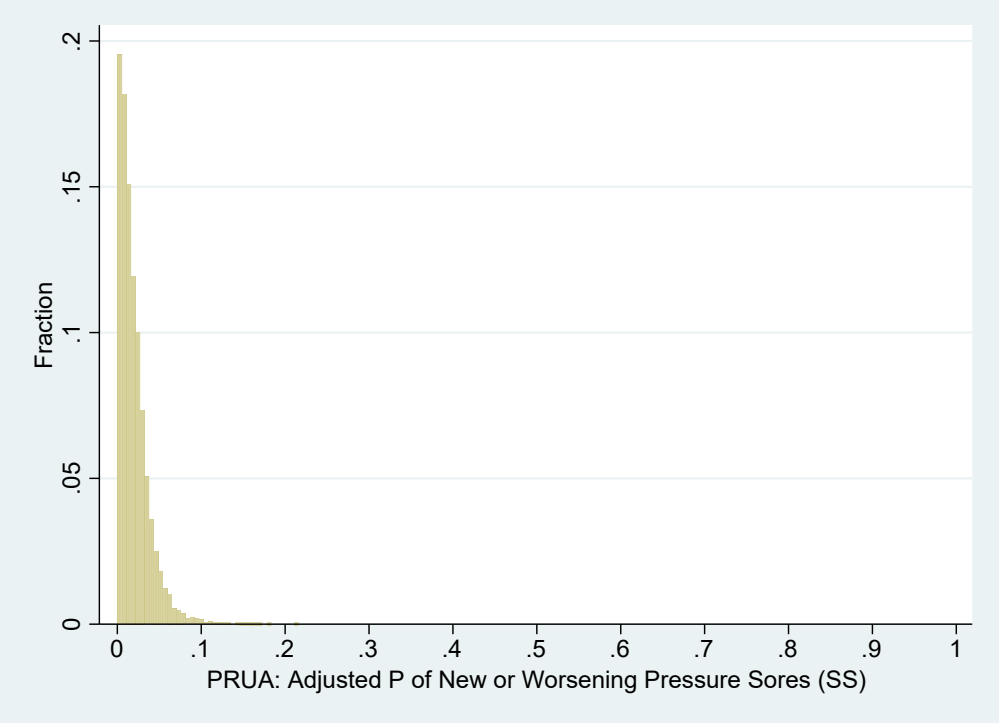
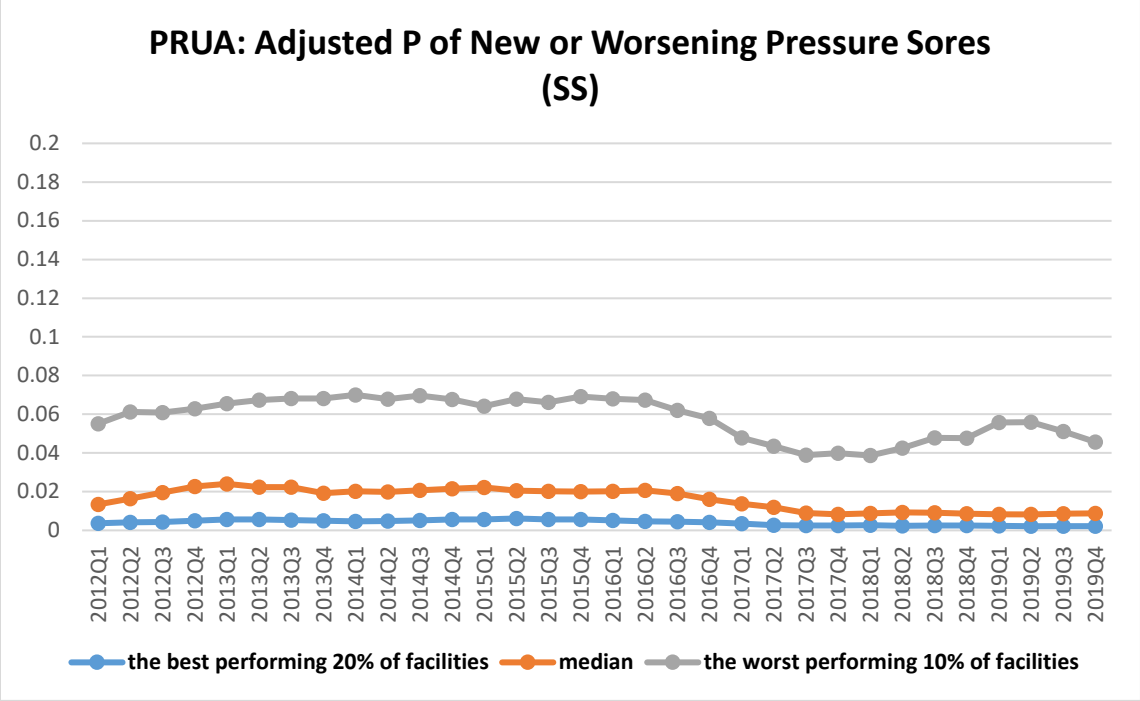


Figure 8: The trends of QI: new or worsening pressure sores



4.3.3 Prevalence of Pressure Sores in High Risk Residents (Long-Stay)

As presented in Figure 9 and Figure 10, the distribution of the QI "prevalence of pressure sores in high risk residents" has a floor effect (mean: 0.039; variance: 0.0004).The trends were stable over time.

Figure 9. The distribution of QI: pressure sores in high risk residents

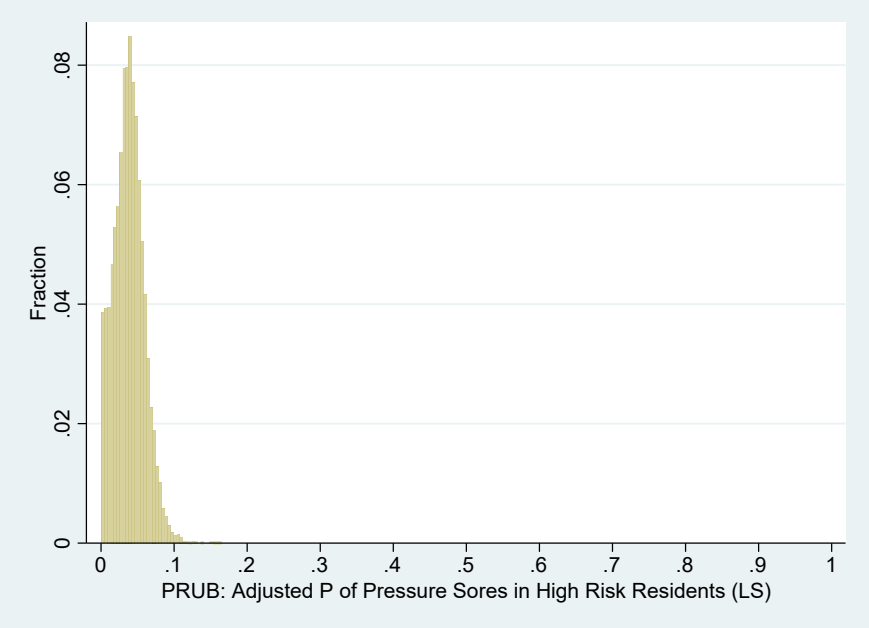
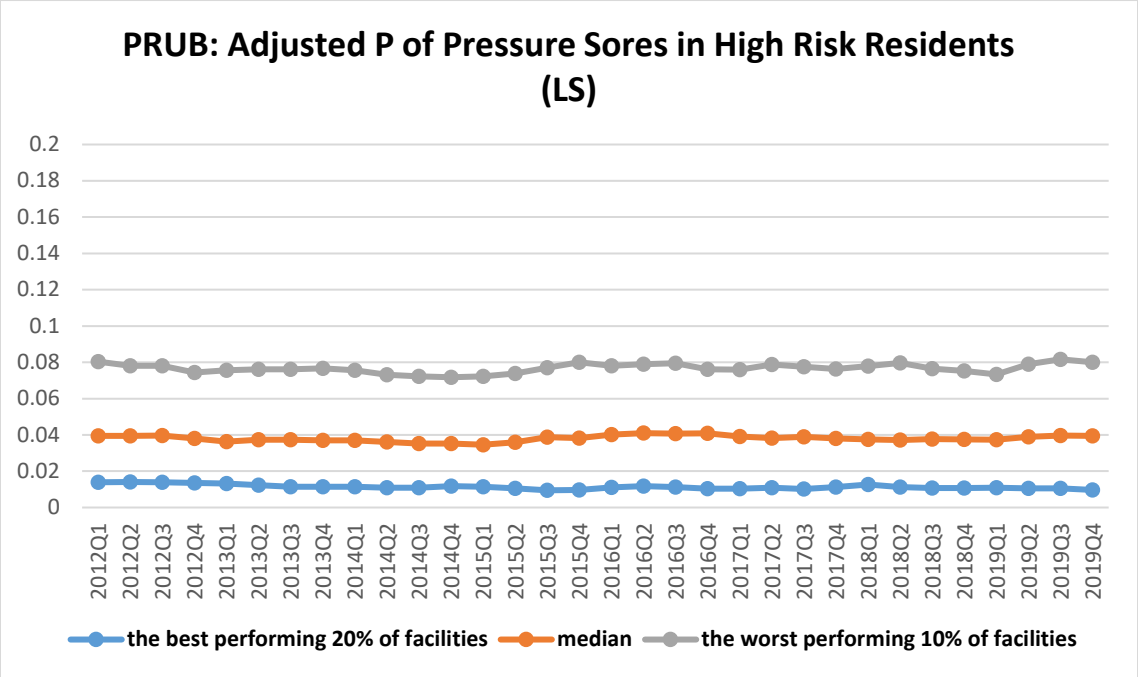


Figure 10: The trends of QI: pressure sores in high risk residents



4.3.4 Prevalence of Indwelling Catheters (Long-Stay)

As presented in Figure 11 and Figure 12, the distribution of the QI "prevalence of indwelling catheters" is skewed with a floor effect (mean: 0.027; variance: 0.0006). The worst performing 10% facilities were improving over time.

Figure 11. The distribution of QI: indwelling catheters

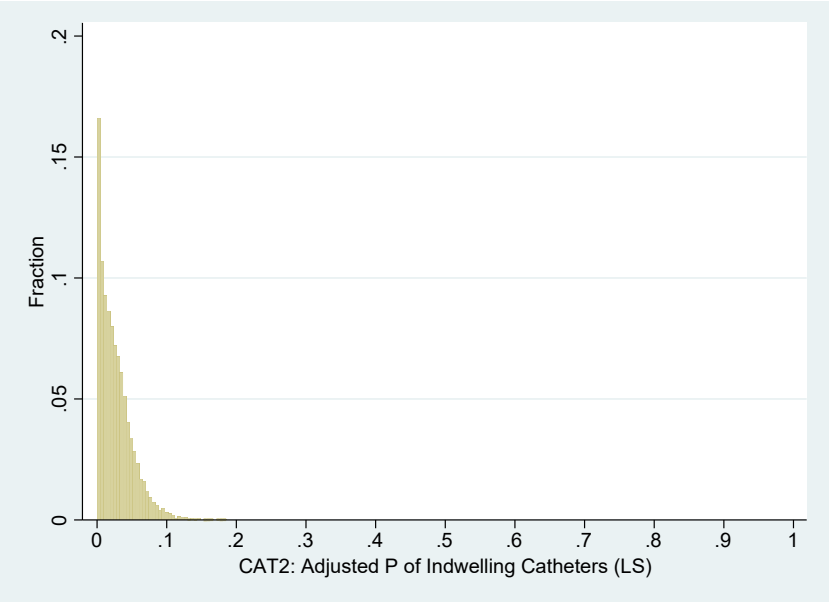
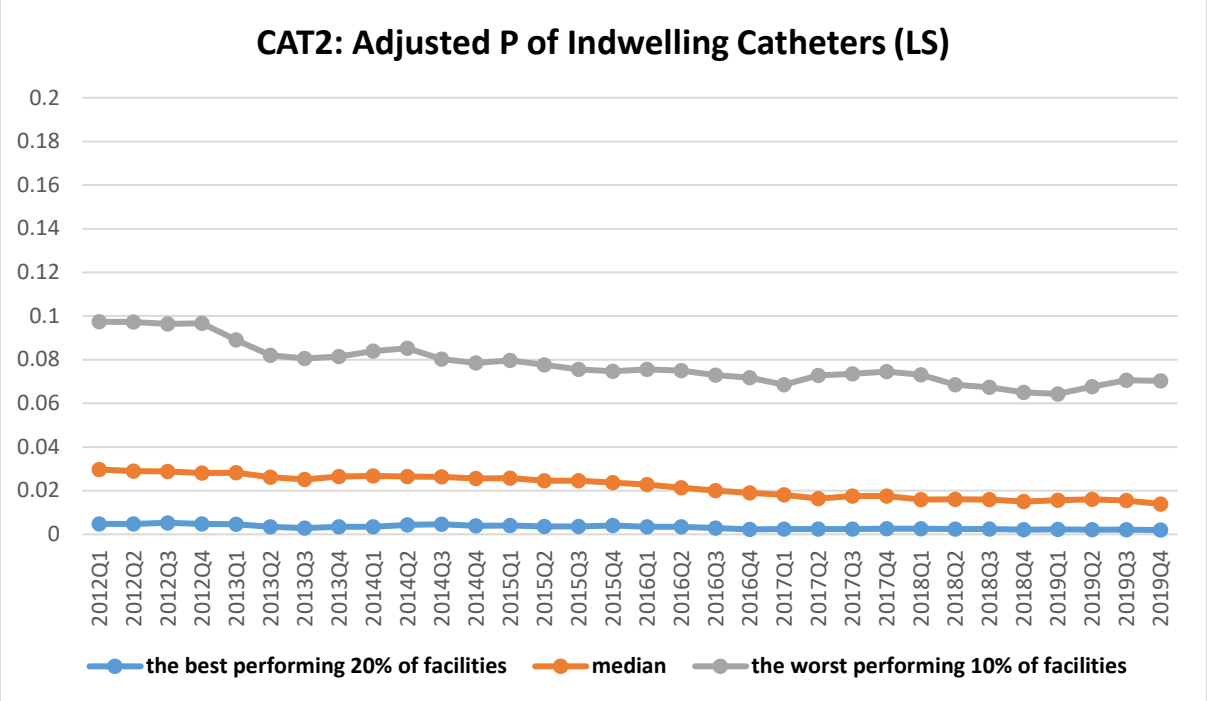


Figure 12: The trends of QI: indwelling catheters



4.3.5 Prevalence of Infections (Long-Stay)

As presented in Figure 13 and Figure 14, the distribution of the QI "prevalence of infections" is skewed with a floor effect (mean: 0.028; variance: 0.0007). The worst performing 10% facilities were improving over time.

Figure 13. The distribution of QI: infections

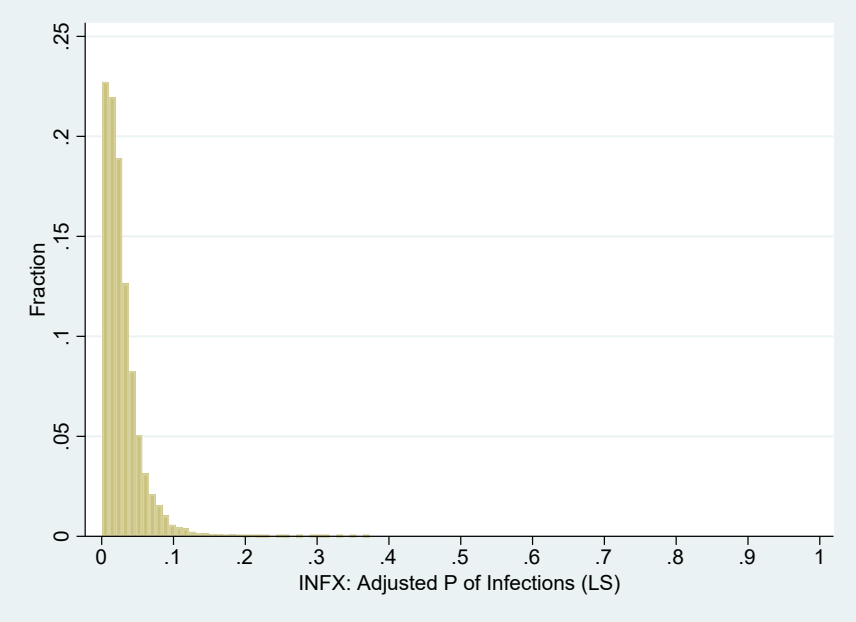
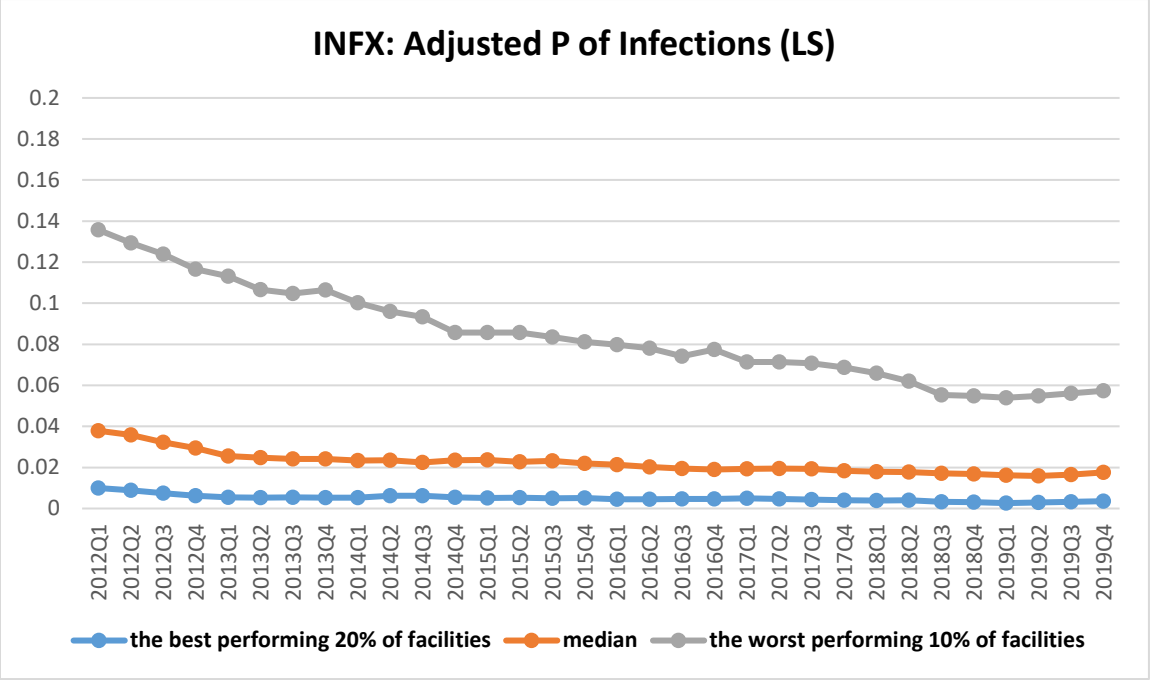


Figure 14: The trends of QI: infections



4.3.6 Prevalence of Falls with Injury (Long-Stay)

As presented in Figure 15 and Figure 16, the distribution of the QI "prevalence of falls with injury" is skewed with a floor effect (mean: 0.034; variance: 0.0007). The trends were stable over time.

Figure 15. The distribution of QI: fall with injury

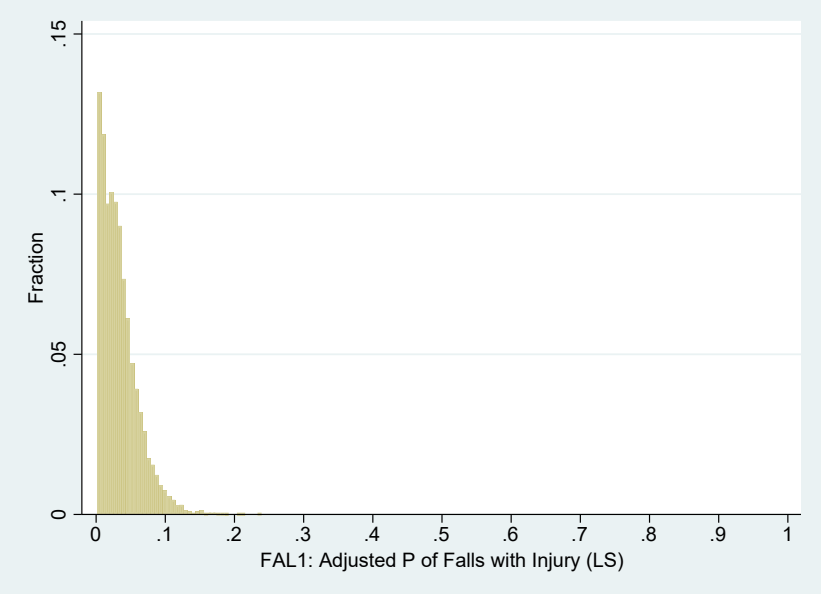
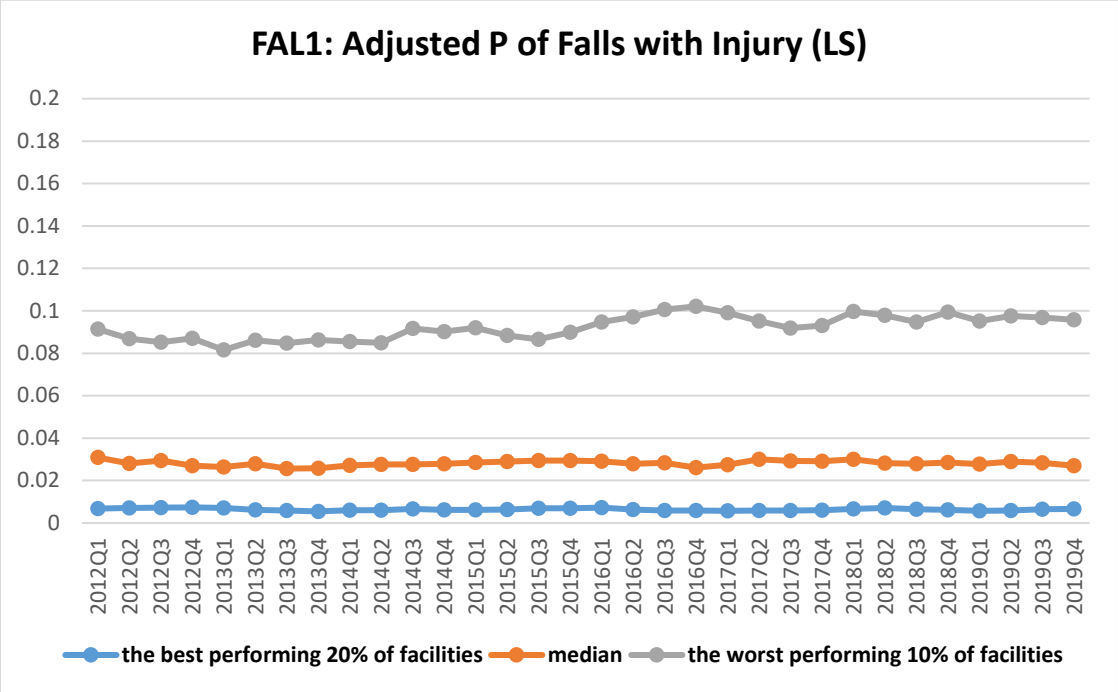


Figure 16: The trends of QI: fall with injury



4.3.7 Prevalence of Urinary Tract Infections (LS)

As presented in Figure 17 and Figure 18, the distribution of the QI "prevalence of urinary tract infections" has a floor effect (mean: 0.041; variance: 0.0009). The worst performing 10% facilities were improving over time.

Figure 17. The distribution of QI: urinary tract infections

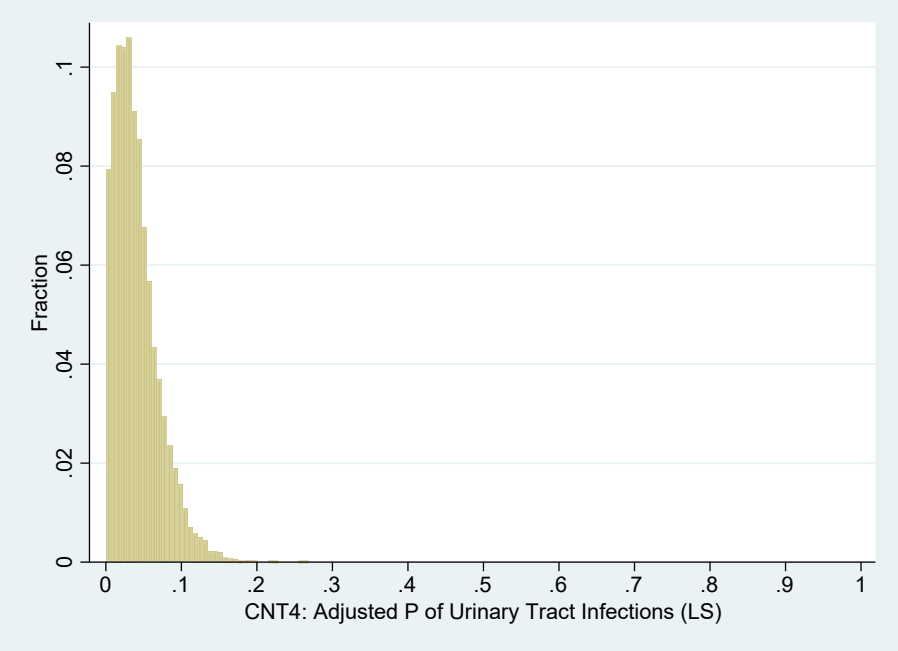
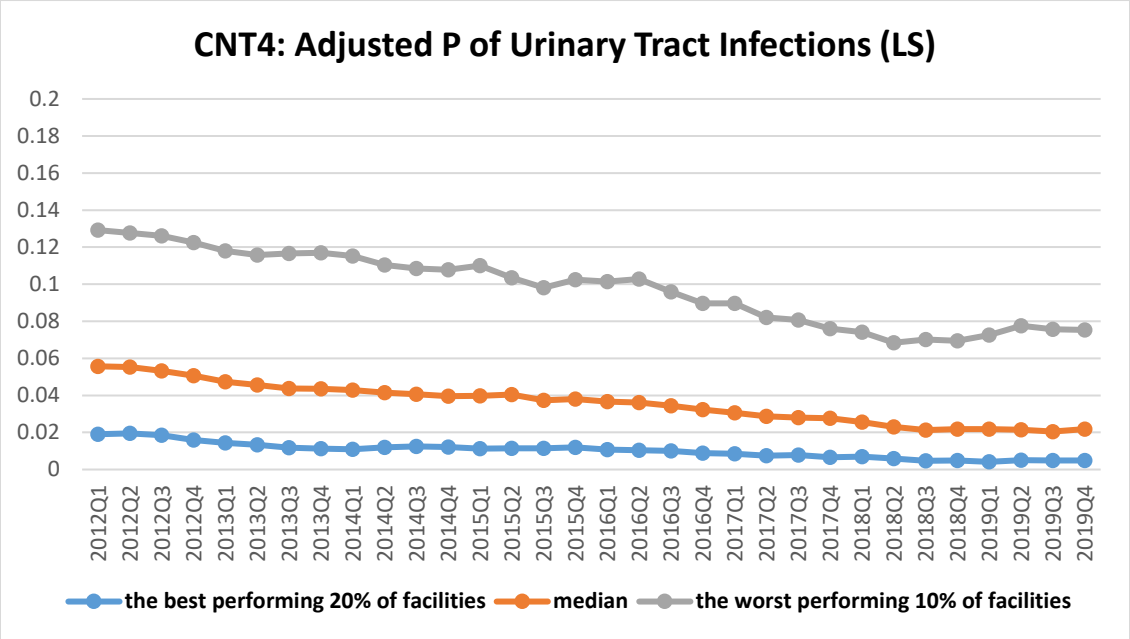


Figure 18: The trends of QI: urinary tract infections



4.3.8 Prevalence of Unexplained Weight Loss (Long-Stay)

As presented in Figure 19 and Figure 20, the distribution of the QI "prevalence of unexplained weight loss" has a floor effect (mean: 0.052; variance: 0.0009).

Figure 19. The distribution of QI: unexplained weight loss

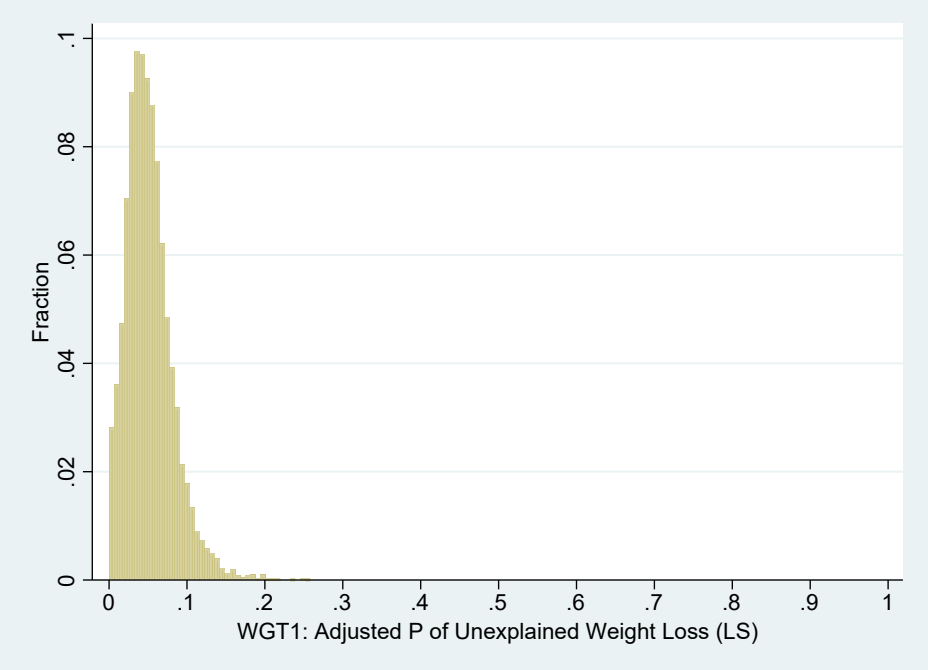
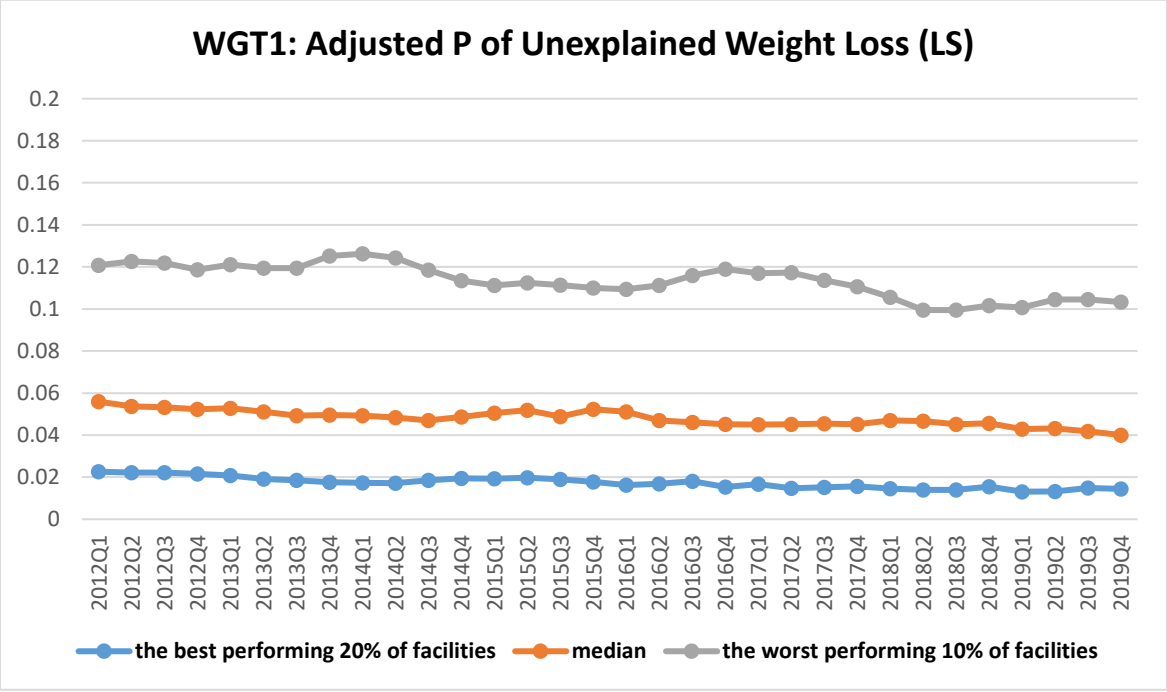


Figure 20: The trends of QI: unexplained weight loss



4.4 Non-Normal Distribution of 4 Long-Stay QIs

There are 4 QIs which do not follow a normal distribution with ceiling or floor effects.

- **Prevalence of occasional to full bladder incontinence without a toileting plan (Figures 21 & 22, ceiling effect)**
- **Prevalence of occasional to full bowel incontinence without a toileting plan (Figures 23 & 24, ceiling effect)**
- **Prevalence of antipsychotic medications without a diagnosis of psychosis (Figures 25 & 26, floor effect)**
- **Prevalence of depressive symptoms (Figures 27 & 28, floor effect)**

The 2 bladder and bowel incontinence without a toileting plan QIs have ceiling effects (Figures 21-24). There are smaller differences between median and the average score of the worst performing 10% of facilities, compared with the differences between median and the average score of the best performing 20% of facilities. However, using the current scoring program, the worst performing 10% of facilities got 0 point, the median got about 1 point, and the best performing 20% of facilities got 2 points (As shown in Table 1, in the continence domain, there are 5 QIs and the total points ranged from 0 to 10 distribute equally. So each QI is assigned a point with a range of 0-2.). Facilities getting 0 point (the worst performing 10% of facilities) are so close to the 1-point threshold (median), they are not very distinguishable. When taking measurement error into account, there is almost no difference in the performance of the 0 and 1-point facilities on the two QIs. Thus, the 2 bladder and bowel incontinence without a toileting plan QIs do not discriminate very well. The point threshold should be re-adjusted, so that the points are a better reflection of facility performance. Moreover, the majorities of facilities did a poor job on these two QIs.

The two QIs "prevalence of antipsychotic medications without a diagnosis of psychosis" and "prevalence of depressive symptoms" have floor effects (Figures 25-28). There are smaller differences between median and the average score of the best performing 20% of facilities, compared with the differences between median and the average score of the worst performing 10% of facilities. The worst performing 10% facilities were improving over time.

Figure 21. The distribution of QI: bladder incontinence without a toileting plan

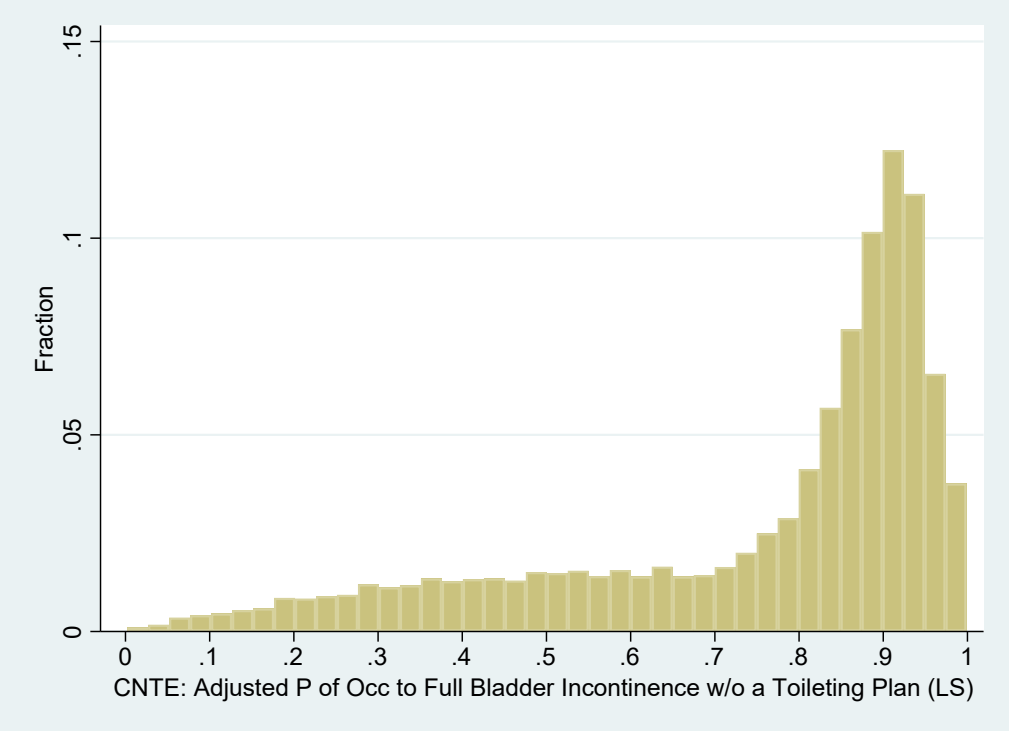


Figure 22. The trends of QI: bladder incontinence without a toileting plan

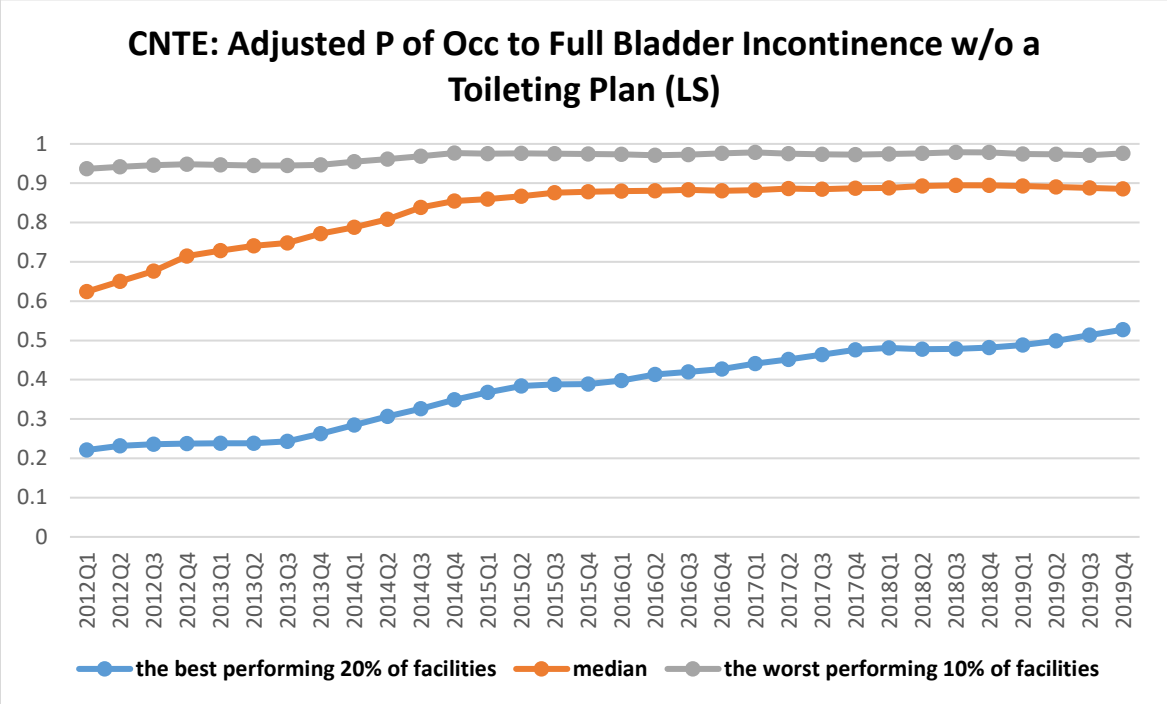


Figure 23. The distribution of QI: bowel incontinence without a toileting plan

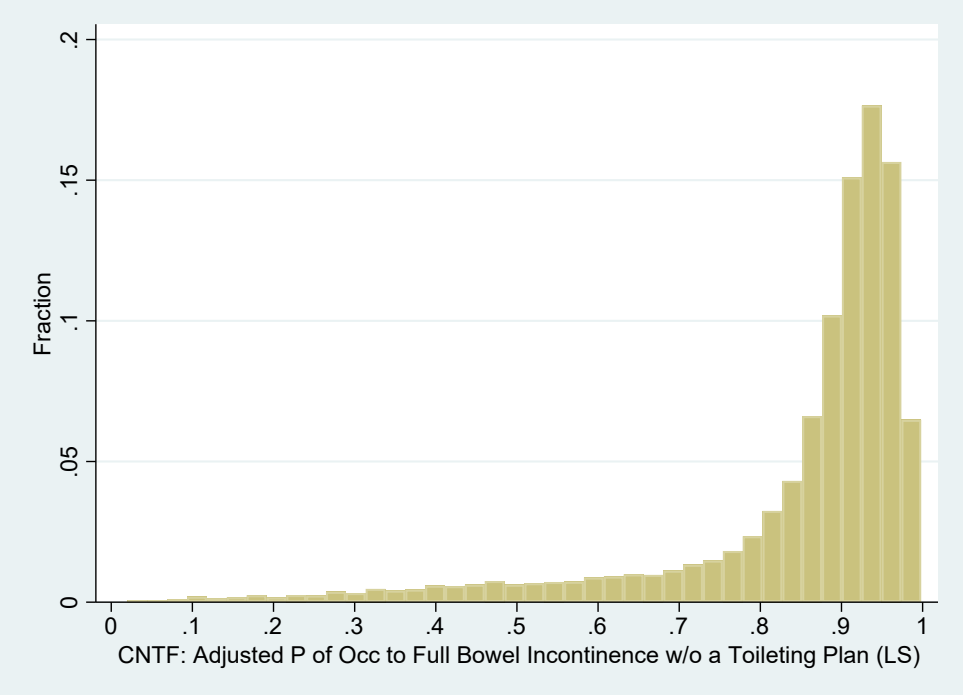


Figure 24. The trends of QI: bowel incontinence without a toileting plan

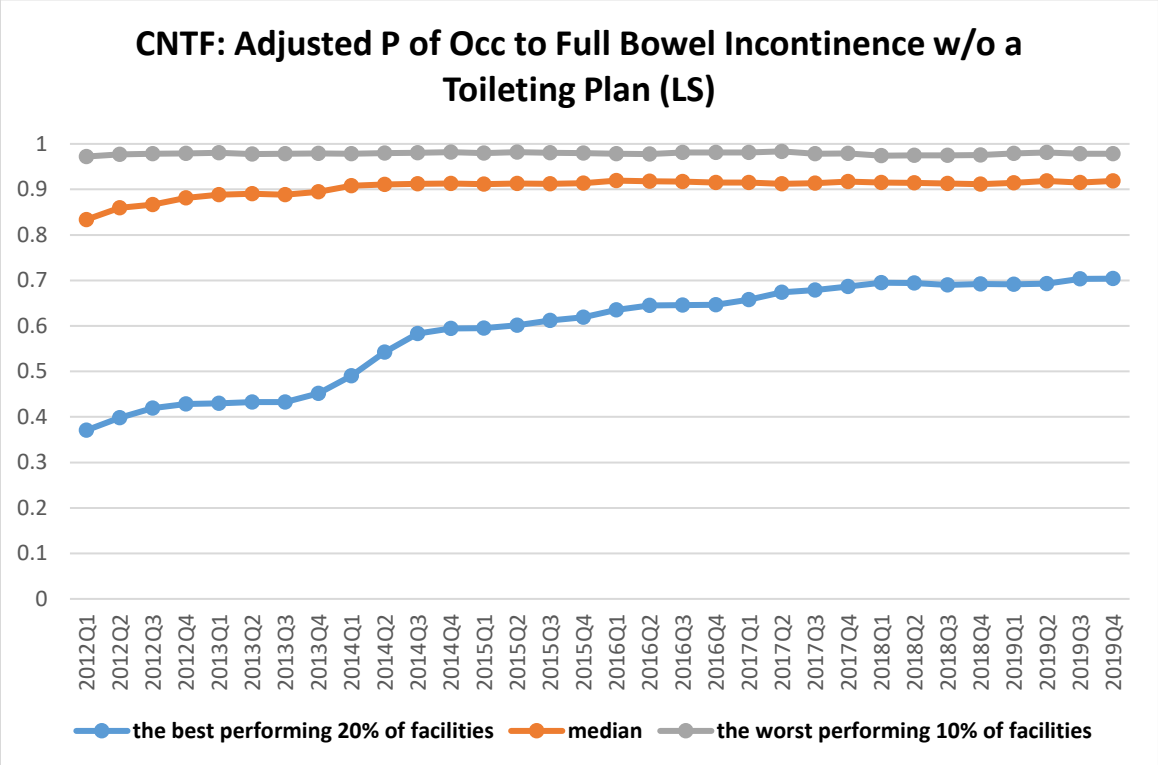


Figure 25. The distribution of QI: antipsychotic medications

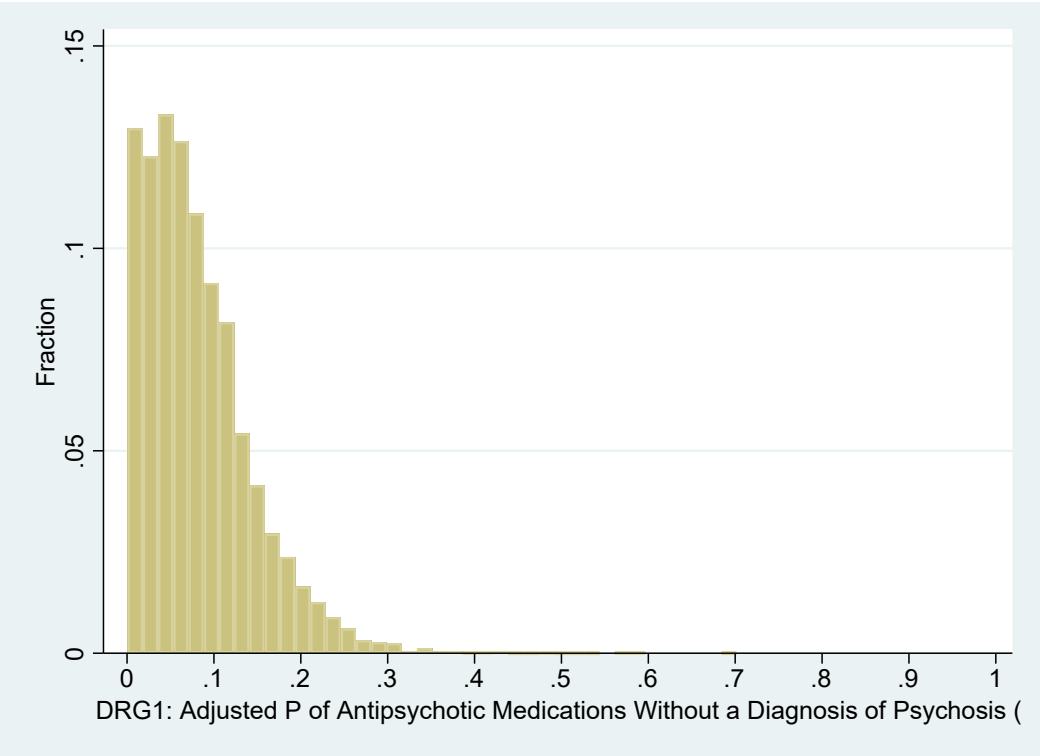


Figure 26. The trends of QI: antipsychotic medications

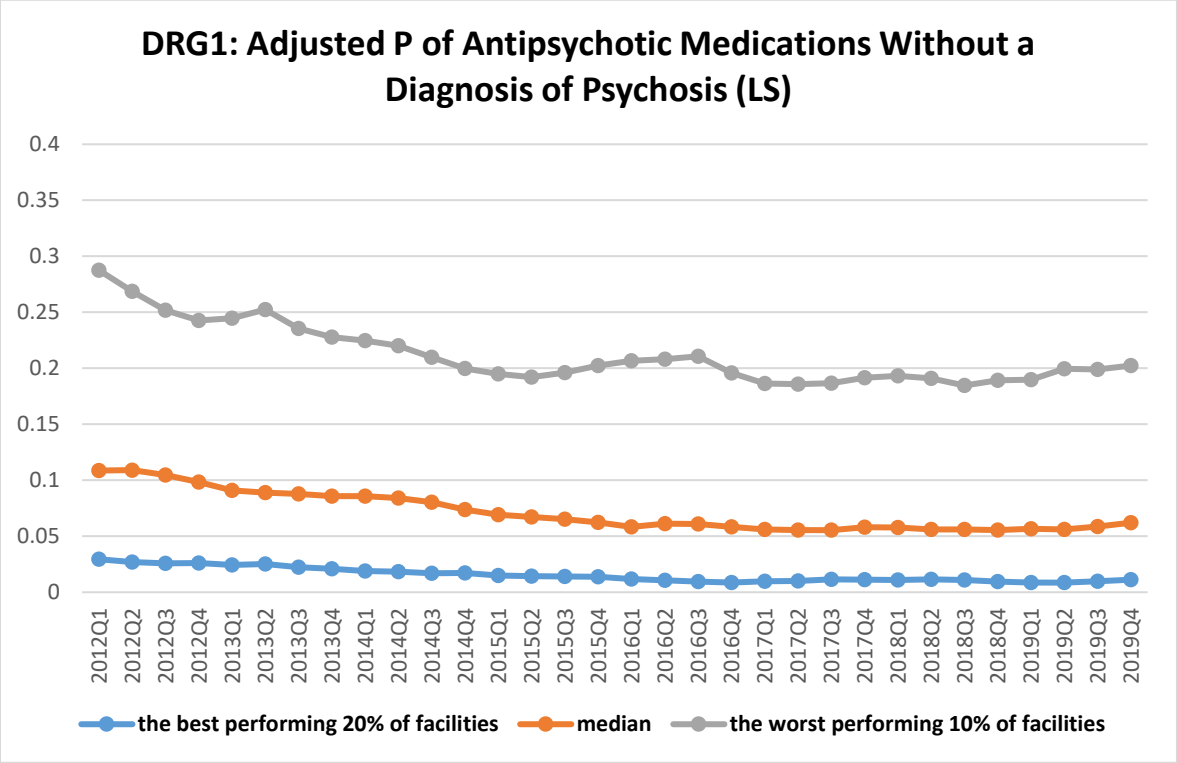


Figure 27. The distribution of QI: depressive symptoms

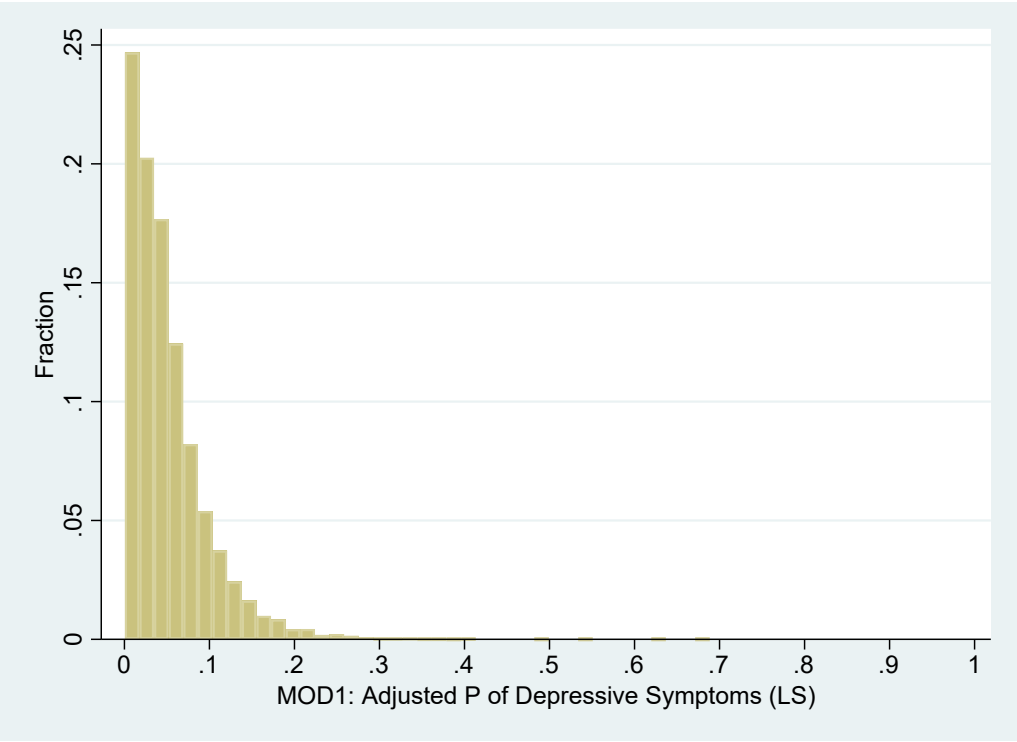
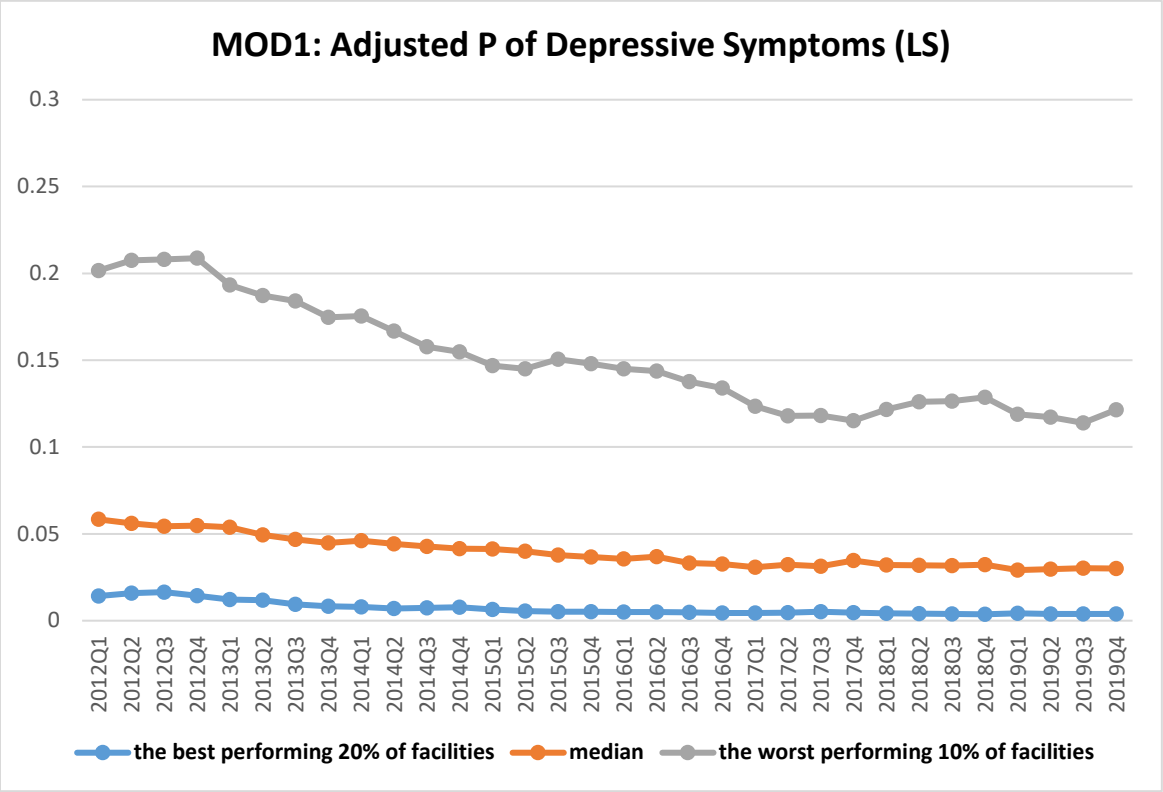


Figure 28. The trends of QI: depressive symptoms



4.5. Approximate Normal Distribution of the Rest 9 QIs

There are 9 QIs which have an approximate normal distribution and considerable variations (Figures 29-46). The current scoring program worked well for these QIs.

- **Incidence of worsening or serious bowel incontinence (long-stay)**
- **Incidence of worsening or serious bladder incontinence (long-stay)**
- **Prevalence of moderate to serious pain (short-stay)**
- **Prevalence of moderate to serious pain (long-stay)**
- **Incidence of worsening or serious mobility dependence (long-stay)**
- **Incidence of worsening or serious functional dependence (long-stay)**
- **Incidence of worsening or serious range of motion limitation (long-stay)**
- **Incidence of worsening or serious resident behavior problems (long-stay)**
- **Incidence of walking as well or better than previous assessment (long-stay)**

Figure 29. The distribution of QI: worsening or serious bowel incontinence

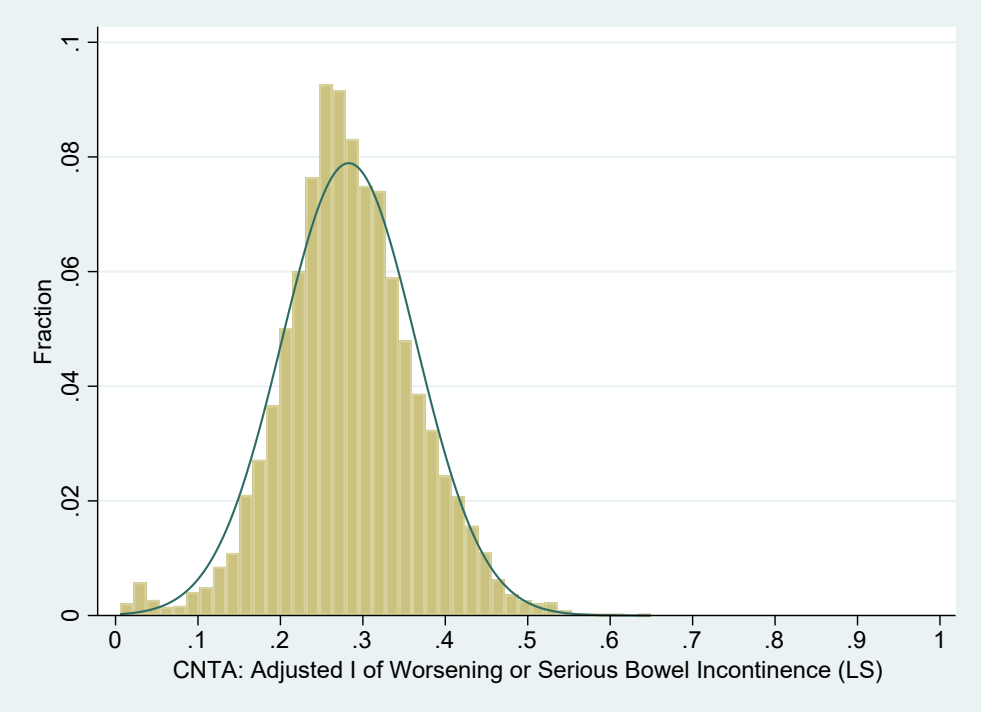


Figure 30. The trends of QI: worsening or serious bowel incontinence

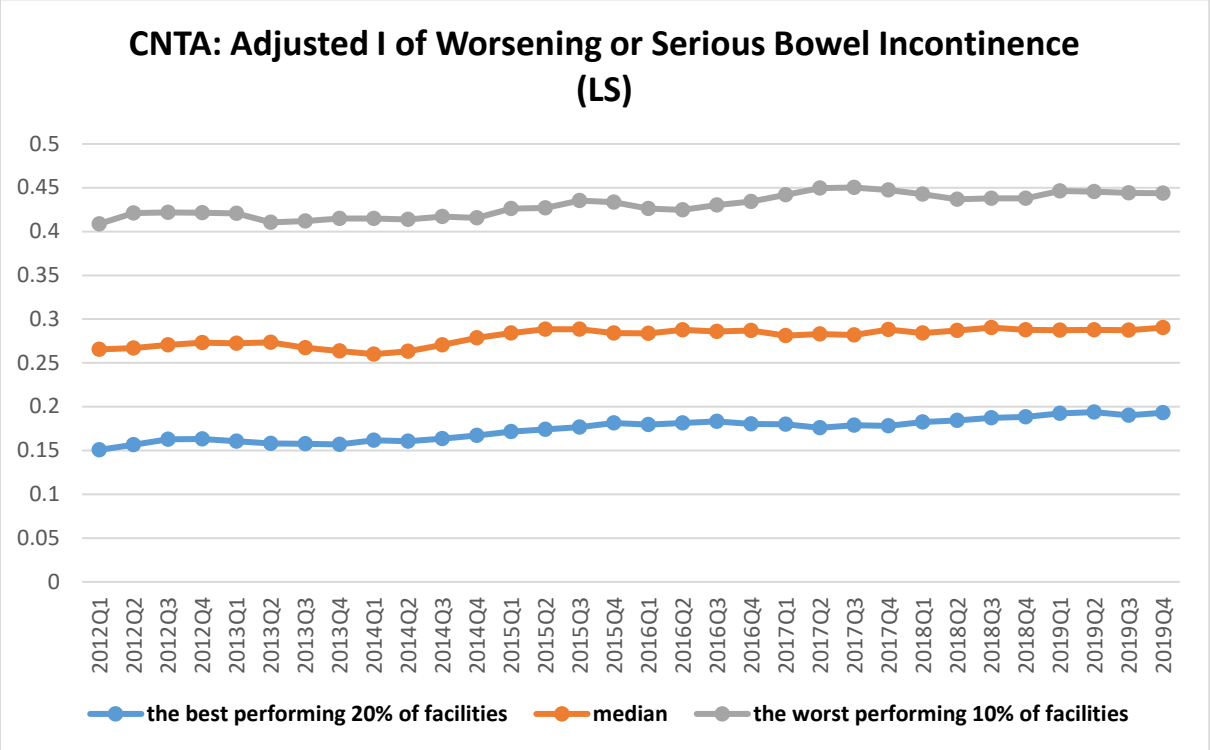


Figure 31. The distribution of QI: worsening or serious bladder incontinence

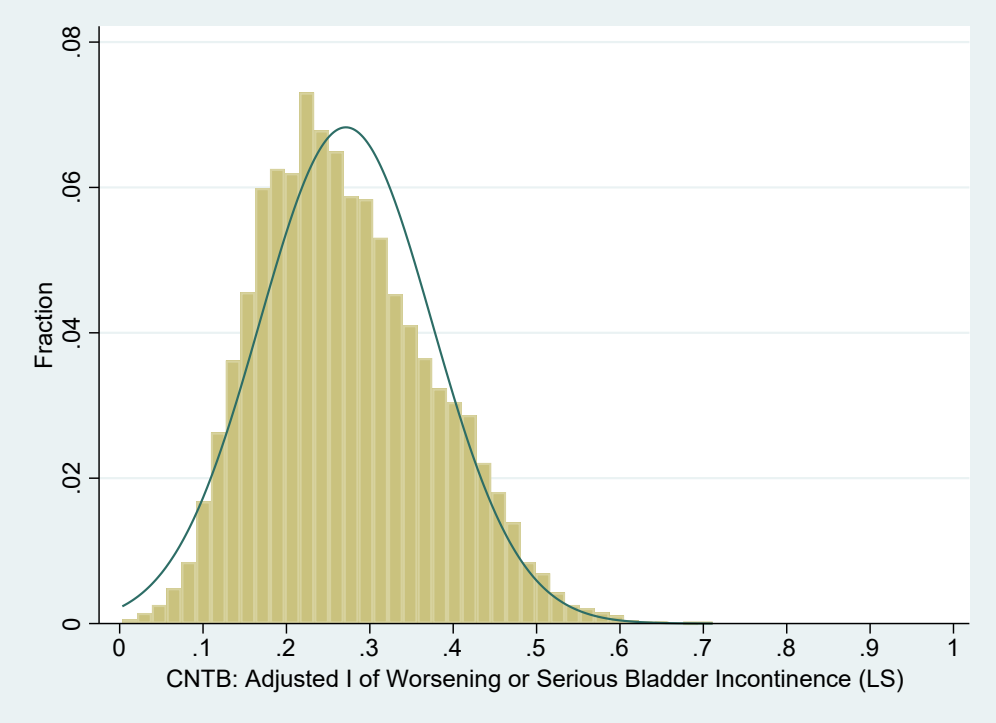


Figure 32. The trends of QI: worsening or serious bladder incontinence

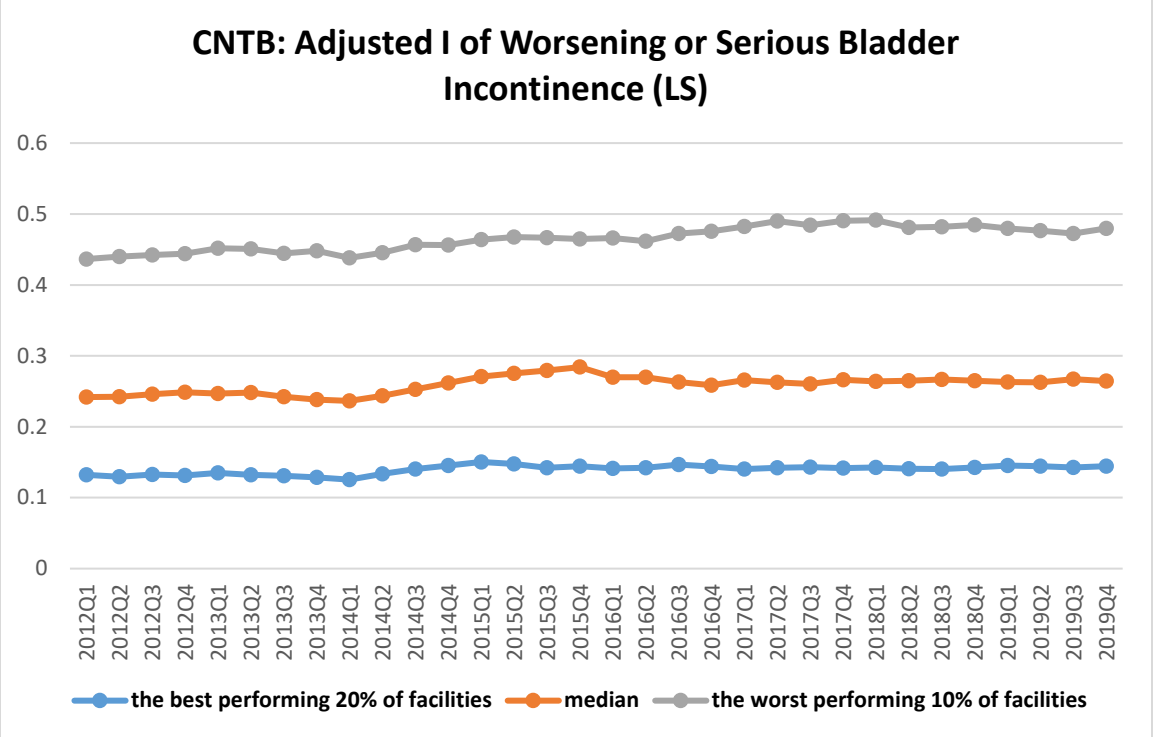


Figure 33. The distribution of QI: short-stay pain

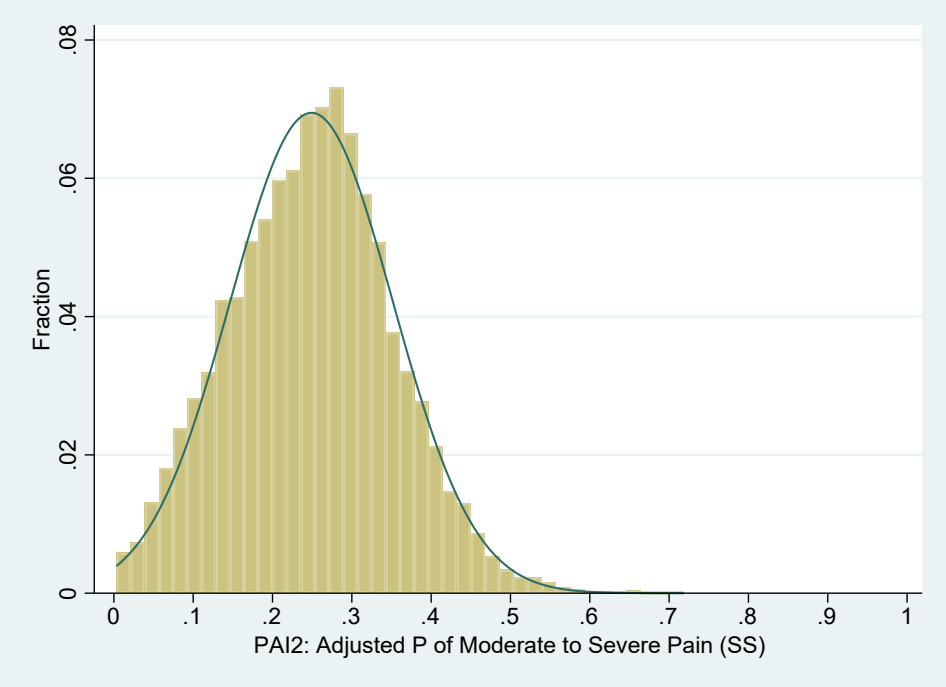


Figure 34. The trends of QI: short-stay pain

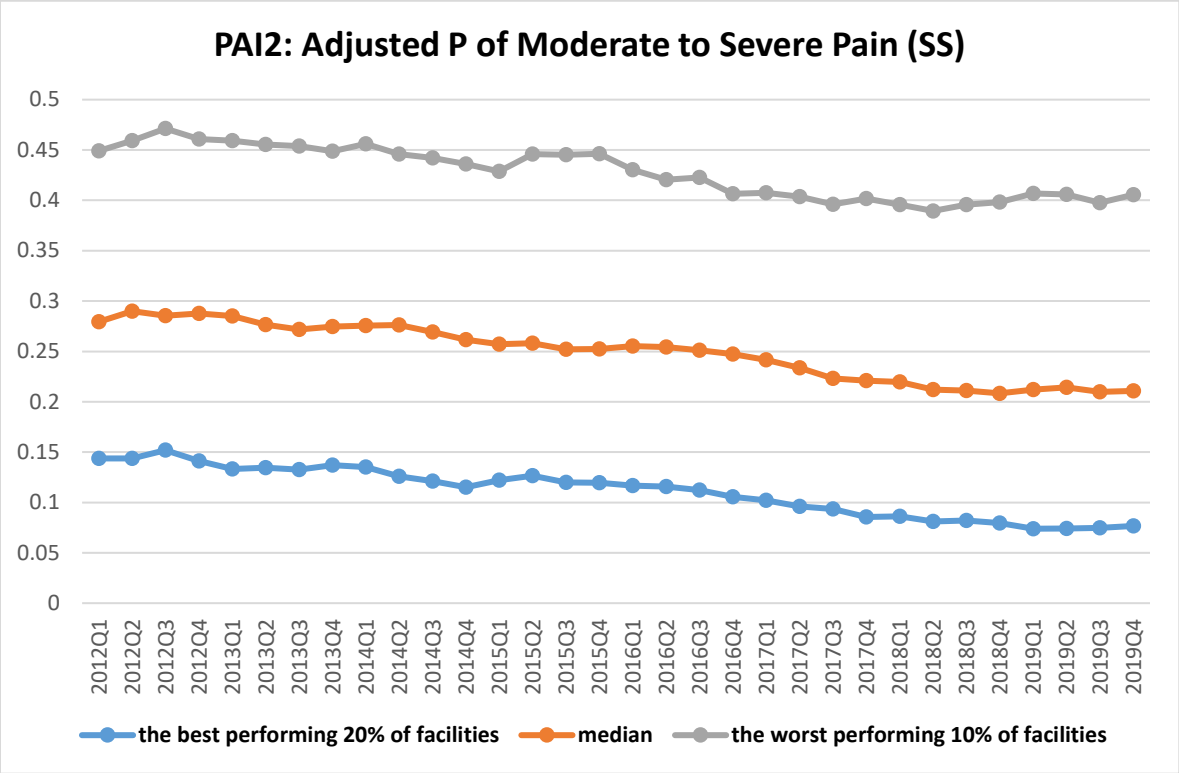


Figure 35. The distribution of QI: long-stay pain

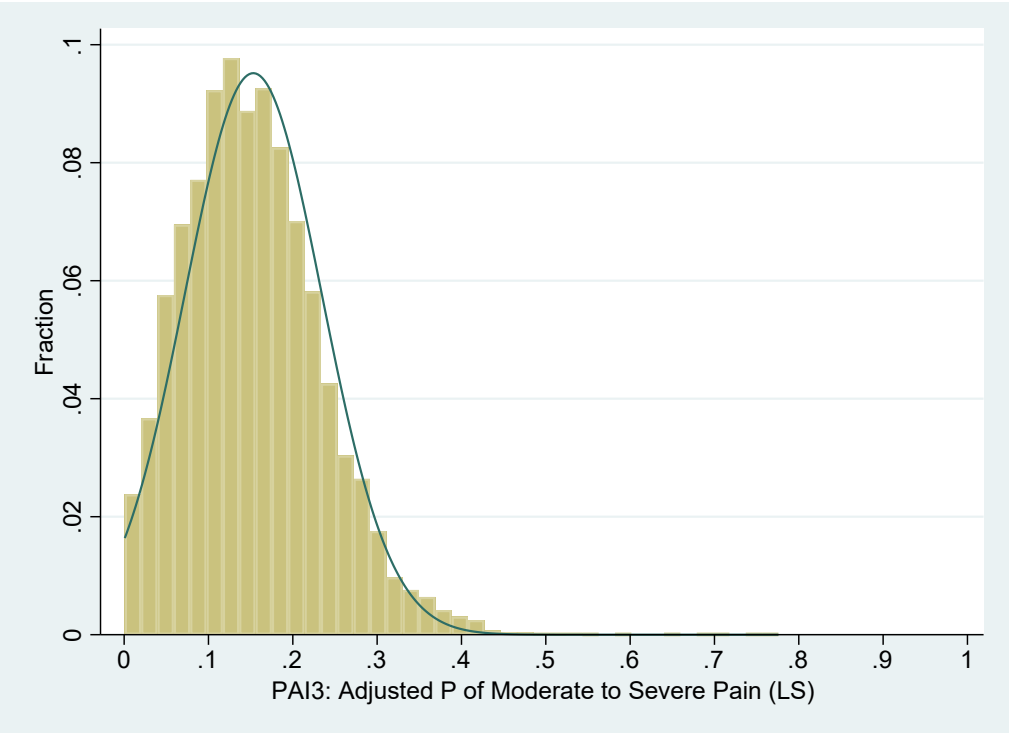


Figure 36. The trends of QI: long-stay pain

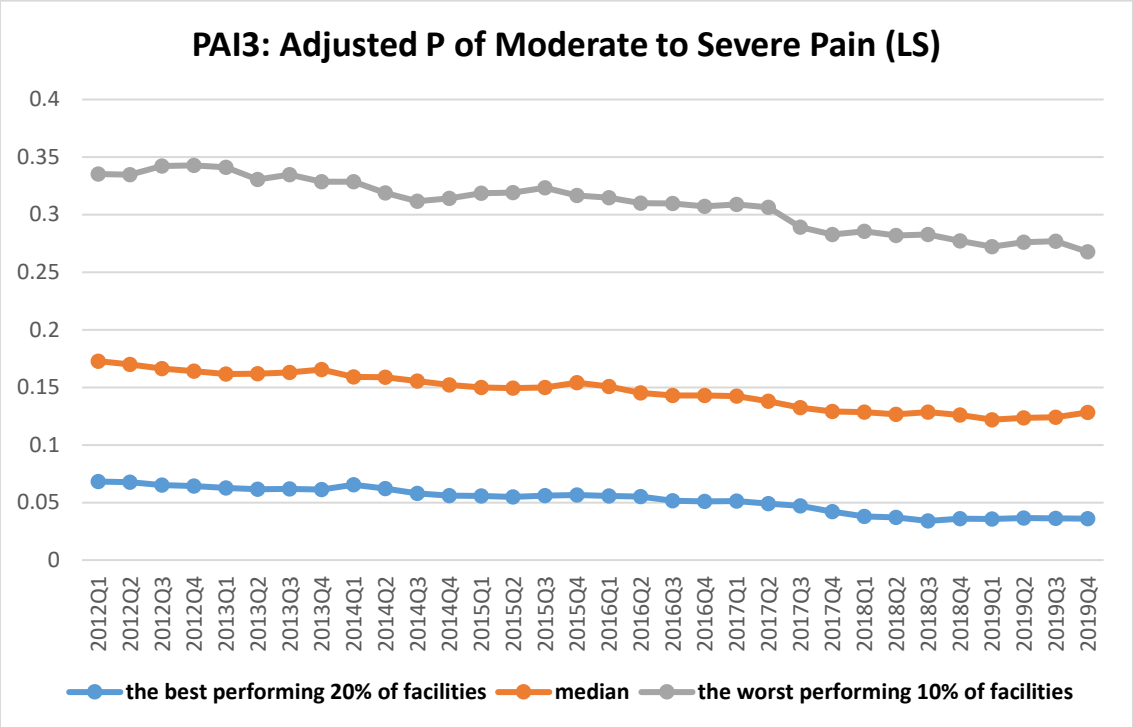


Figure 37. The distribution of QI: mobility dependence

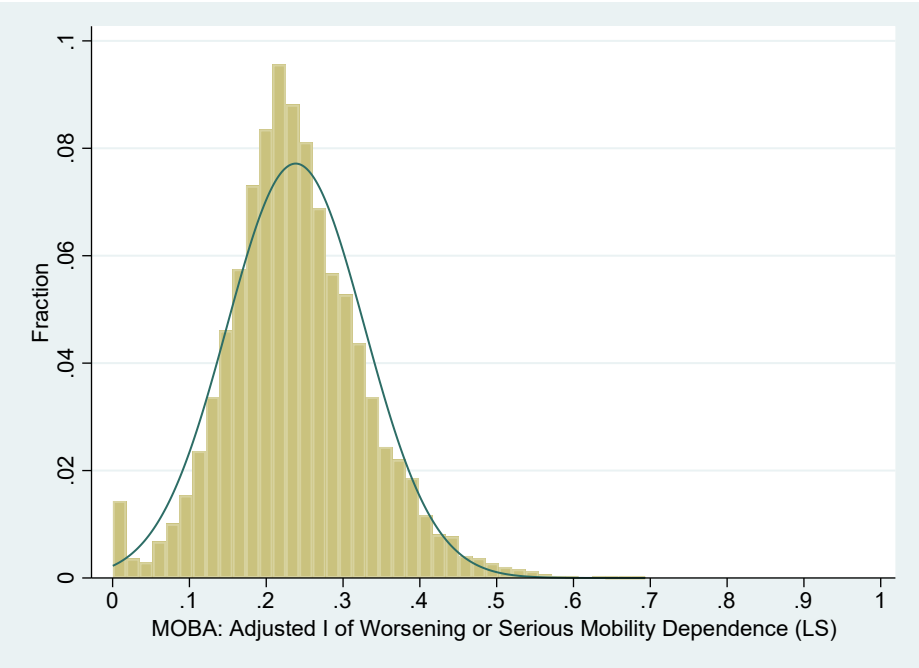


Figure 38. The trends of QI: mobility dependence

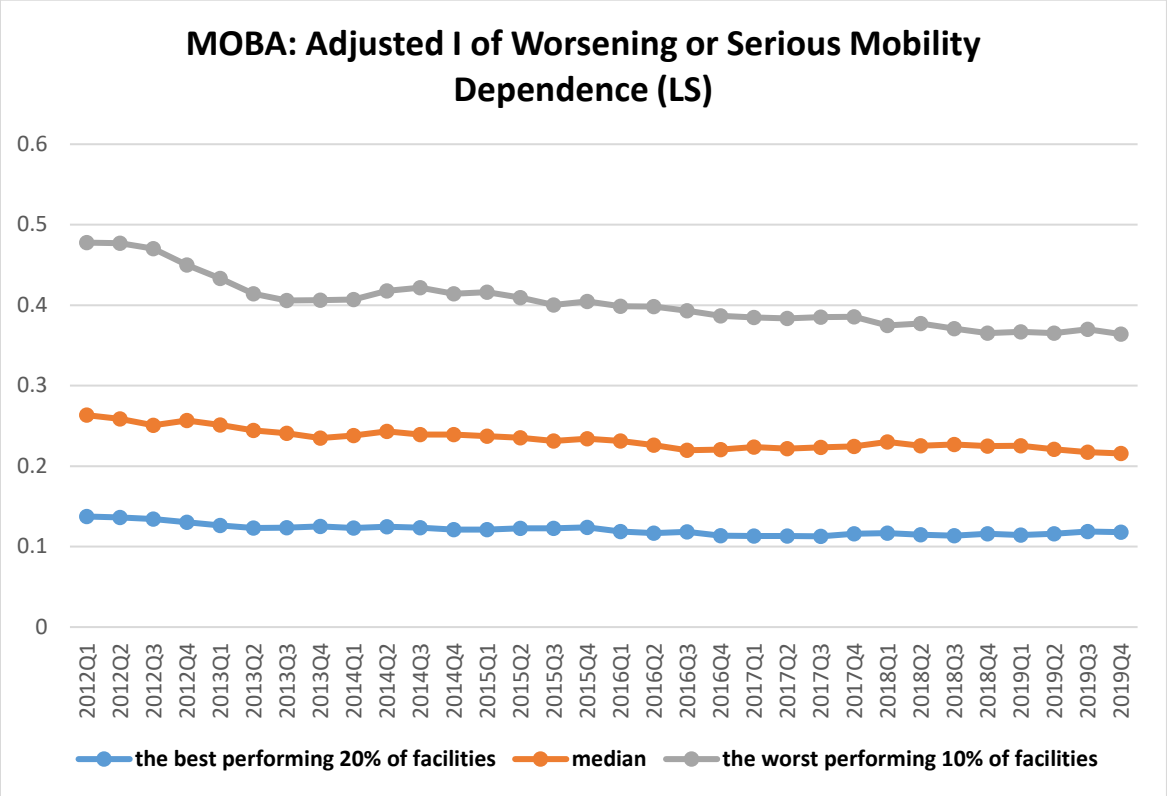


Figure 39. The distribution of QI: functional dependence

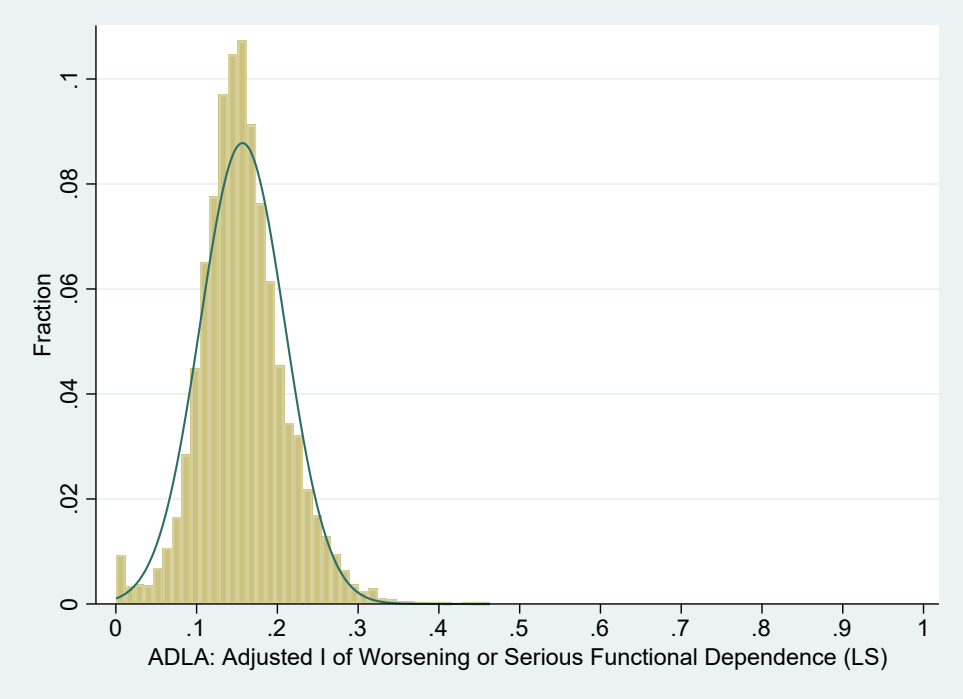


Figure 40. The trends of QI: functional dependence

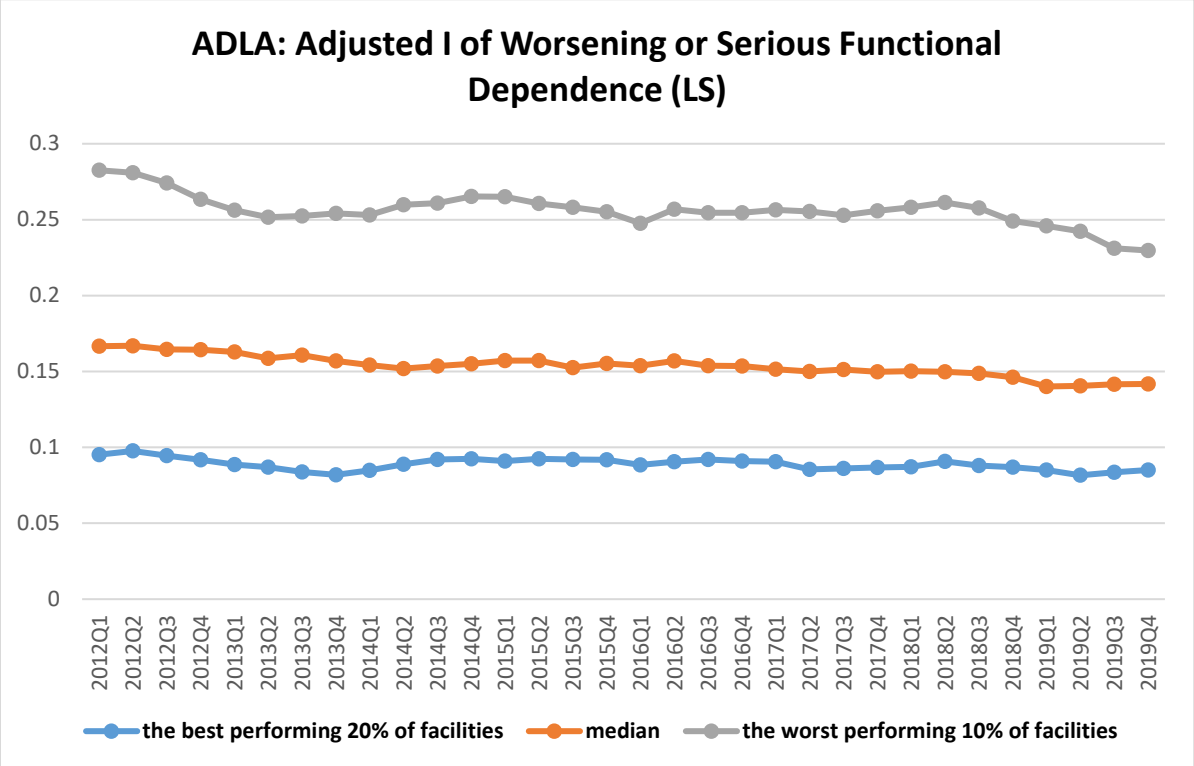


Figure 41. The distribution of QI: range of motion limitation

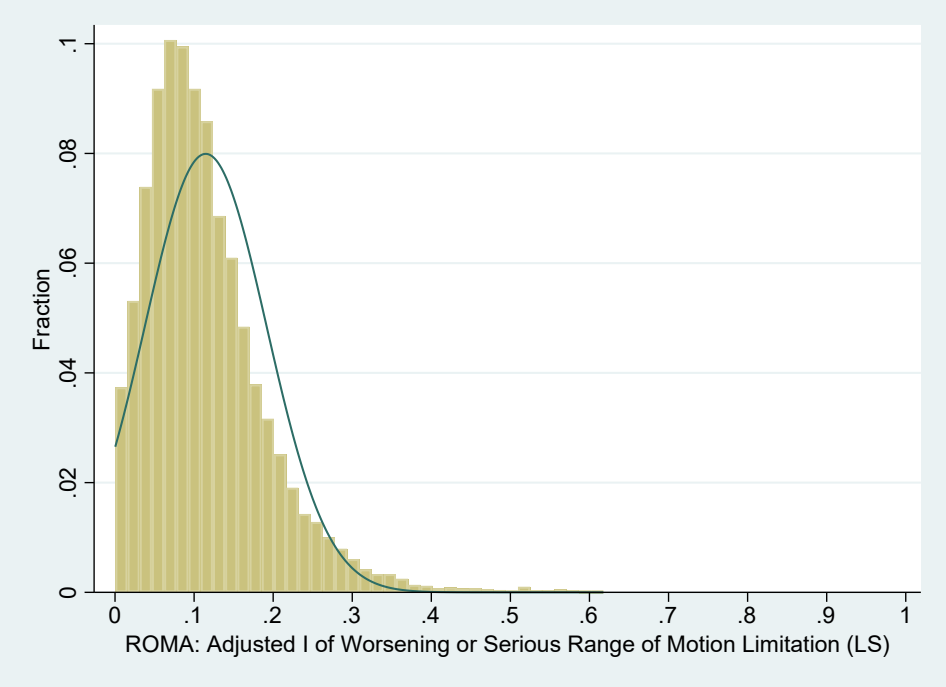


Figure 42. The trends of QI: range of motion limitation

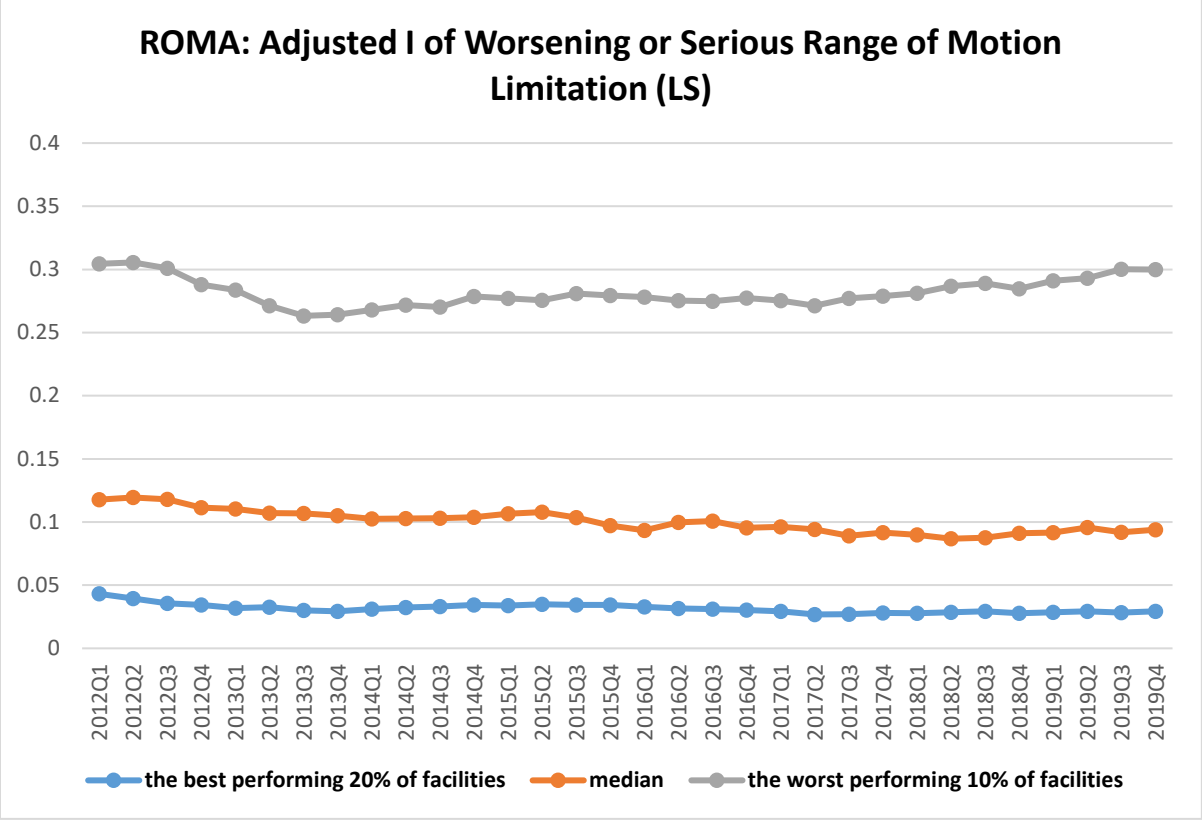


Figure 43. The distribution of QI: resident behavior problems

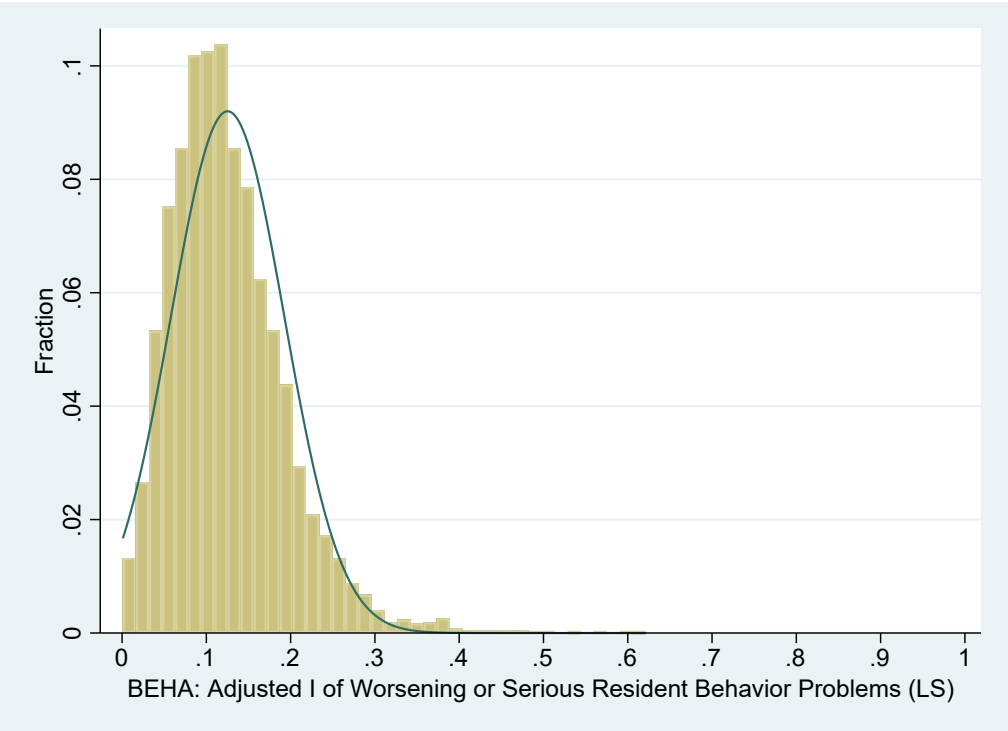


Figure 44. The trends of QI: resident behavior problems

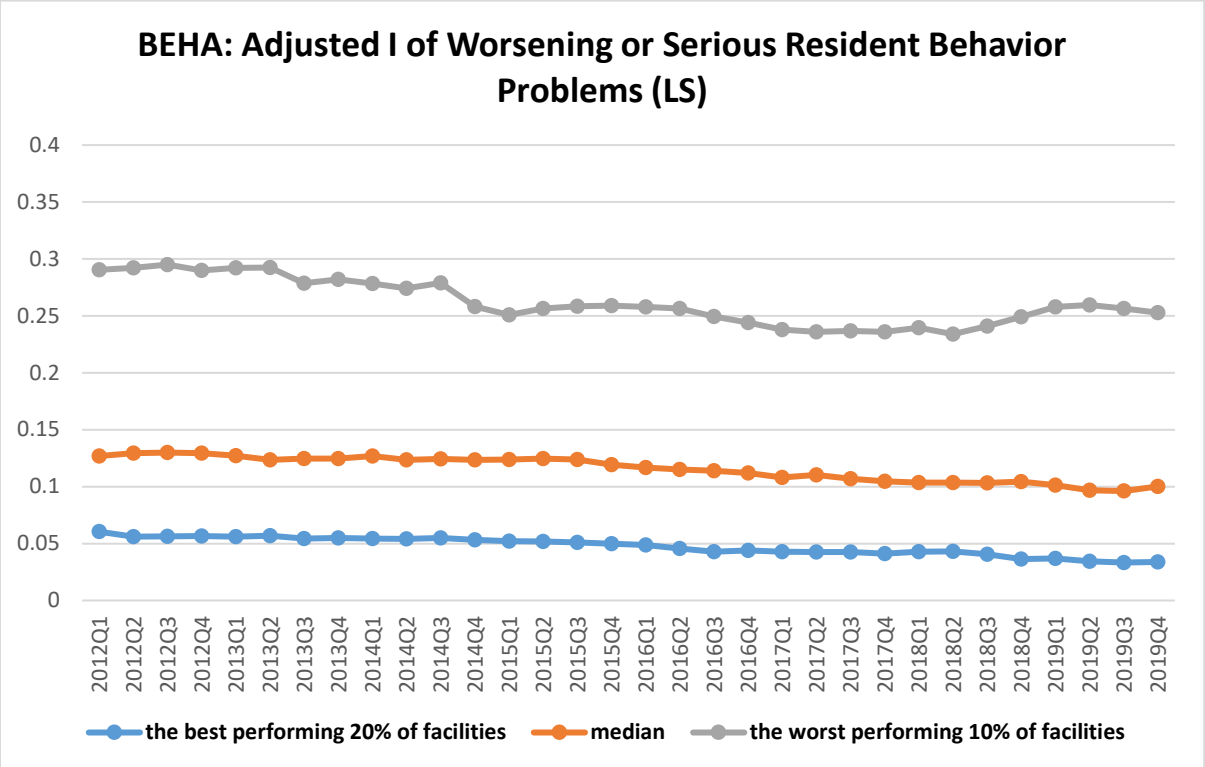


Figure 45. The distribution of QI: walking as well or better

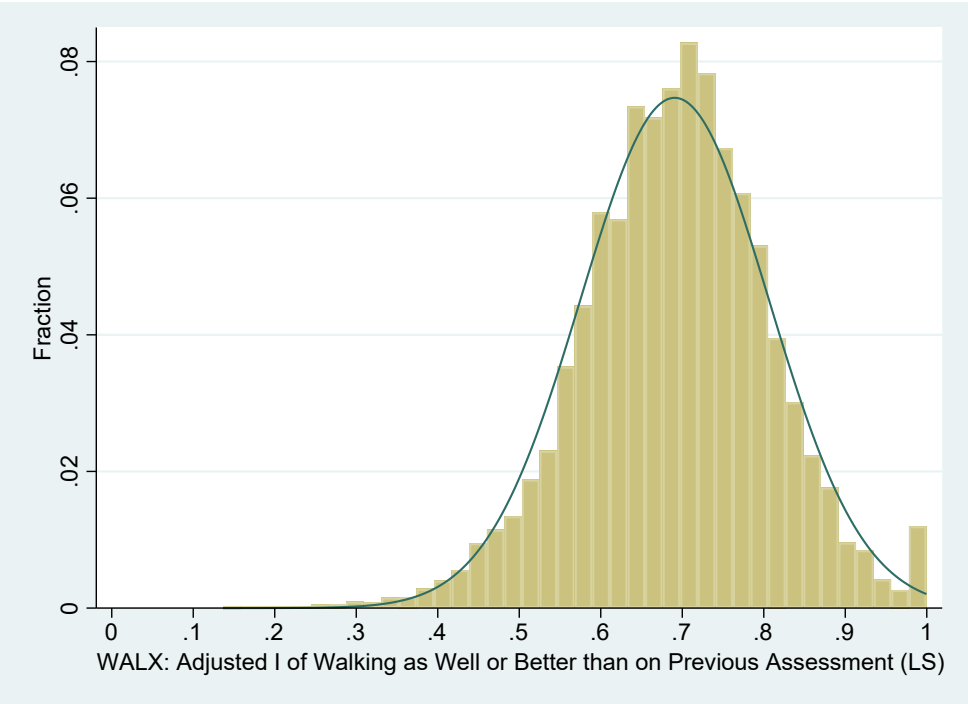
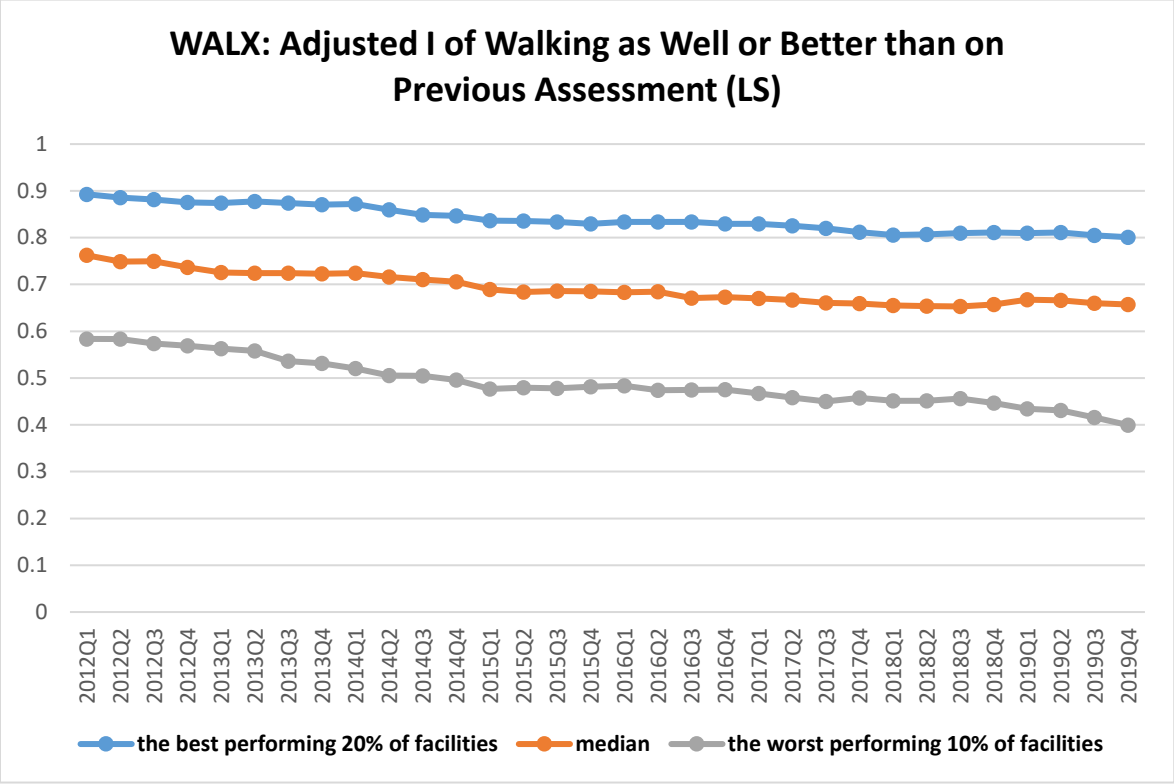


Figure 46. The trends of QI: walking as well or better



5. Summary

Table 7. Summary of 21 QIs

21 Quality Indicators	Factor	Problems	Suggestions
I of Worsening or Serious Bowel Incontinence (LS)	Incontinence	moderate correlation	Combine with bladder continence
I of Worsening or Serious Bladder Incontinence (LS)		moderate correlation	Combine with bowel continence
P of Occasional to Full Bladder Incontinence w/o a Toileting Plan (LS)	Incontinence without a toileting plan	ceiling effect, moderate correlation	Combine with bowel continence w/o a toileting plan, reform the scoring program
P of Occasional to Full Bowel Incontinence w/o a Toileting Plan (LS)		ceiling effect, moderate correlation	Combine with bladder continence w/o a toileting plan, reform the scoring program
P of Moderate to Severe Pain (LS)	Pain		
P of Moderate to Severe Pain (SS)			
I of Walking as Well or Better than on Previous Assessment (LS)	Physical Functioning	the only positive QI	It might be redefined negatively as worsening walking
I of Worsening or Serious Functional Dependence (LS)		correlation	Combine with mobility dependence
I of Worsening or Serious Mobility Dependence (LS)		correlation	Combine with functional dependence
P of Falls with Injury (LS)		too little variance, floor effect	Consider practical significance and reform the scoring program
I of Worsening or Serious Range of Motion Limitation (LS)			
I of Worsening or Serious Resident Behavior Problems (LS)	Restraints and Behavioral Symptoms		
P of Depressive Symptoms (LS)		floor effect	reform the scoring program

21 Quality Indicators	Factor	Problems	Suggestions
P of Physical Restraints (LS)		too little variance, floor effect	Using a threshold value
P of Antipsychotic Medications Without a Diagnosis of Psychosis (LS)		floor effect	reform the scoring program
P of Infections (LS)		too little variance, floor effect	Consider practical significance and reform the scoring program
P of New or Worsening Pressure Sores (SS)		too little variance, floor effect	Consider practical significance and reform the scoring program
P of Pressure Sores in High Risk Residents (LS)		too little variance, floor effect	Consider practical significance and reform the scoring program
P of Unexplained Weight Loss (LS)	Care for Specific Conditions	too little variance, floor effect	Consider practical significance and reform the scoring program
P of Indwelling Catheters (LS)		too little variance, floor effect	Consider practical significance and reform the scoring program
P of Urinary Tract Infections (LS)		too little variance, floor effect	Consider practical significance and reform the scoring program

Besides the summary in Table 7, there are additional two things worth discussing. First, the line graphs also paint a picture of change in the QIs over time. Some of the change may be due to a change in the MDS or the way the QIs are defined. Some of the trends in the QIs may indicate changes in true care quality. For example, restraints and infections tend to be trending downward, while absence of a toileting plan is trending upward. Second, since the thresholds are based on percentiles, they will move with overall QI trends. A specific QI rate (10% incidence or prevalence) could move up or down in the point system over time depending on how the distribution changes. Conversely, a facility may be improving in its QI rate, but since others are improving as well, that facility will not get any higher points. Should the thresholds possibly be fixed, that is based on the same QI rates over time? This is a potential area for future analysis.

6. Next Steps

- **Exploratory factor analysis using resident-level QIs to determine whether aggregation at the resident- and facility-level QIs yields the same underlying dimensions or domains.**
- **Explore the distribution of points assigned for each QI.**

- **Examine the relationships between QI and overall QI score and other quality measures.**
- **Explore the possibility of changing the weights assigned to QIs.**

Chapter 3

Literature Review: Value-Based Reimbursement and NH Quality

Kathleen Abrahamson, PhD RN

Executive Summary

A search of the academic literature was completed to capture recent evidence surrounding factors that influence quality measurement and the dimensions in the relationship between value base purchasing (VBP) and quality. Thirty-six research articles were identified and described in this report. Evidence fell into 3 broad categories: nursing home quality, reimbursement and costs; influence of VBP on care quality and outcomes; and the nursing home report card and MDS quality measures. Search results are presented in the form of synthesized key findings, a summary of study findings organized by theme, and a table to provide an overview of individual studies.

Introduction

This report describes a search and review of academic literature that was completed to identify and summarize recent research regarding the relationships between value based purchasing (VBP) reimbursement policy, quality measurement, facility spending decisions, and care quality outcomes within nursing homes. The search addressed the following questions:

1) **What factors influence quality measurement in nursing homes (NHs)?**
2) **What is the influence of VBP programs on care quality outcomes?**

Search Methods

A search of the PubMed database was completed using the search terms "nursing home" (MeSH for Skilled Nursing Facility (SNF)) AND quality AND measurement, resulting in 753 research articles. The search was limited to research published within the past 5 years in order to highlight recent additions to the literature, and included only publications from within peer-reviewed journals. Research was excluded from this summary if it addressed quality measurement in a setting apart from nursing homes/ SNFs; did not directly address factors that influence quality outcomes; or did not address the relationship between resources and quality. Abstracts were reviewed and 21 applicable papers were obtained. An additional 15 articles were identified through an ancestry search of the reference list of identified articles. Articles identified through the ancestry search were allowed to go beyond the 5 year date limit in an effort to capture frequently cited and pivotal works in this area. A total of 36 primary research articles from peer reviewed journals contributed to this summary. Unsurprisingly, there was wide variation in identified studies given the broad nature of the search and minimal exclusion criteria.

Search Results

Identified studies ranged in publication date from 1998-2019, and came from a wide variety of high quality nursing, gerontology, medical, economics and health services journals. Studies were most commonly retrospective analyses of large government databases such as the MDS, OSCAR, Medicare Claims Data, and the Area Resource File, with the exception of 4 surveys, 3 commentaries, 3 interview based studies, 1 systematic review and 1 mixed methods study combining secondary data analysis with observations of care.

Key takeaways from the synthesized findings:

1. NH quality, reimbursement, and costs

- **Increased reimbursement does not necessarily correlate with improvements in quality.**
- **Financial constraints are not clearly predictive of the inability to deliver quality care, and processes unmeasured by quality indicators (QI) such as leadership stability and team approaches to care may play a larger role in quality than spending.**
- **The relationship between costs and quality is variable and often inverse. High cost were sometimes correlated with high quality, but often low costs were correlated with high quality. High costs were often correlated with low quality, highlighting the costs of managing the outcomes of poor quality care such as falls and pressure ulcers.**
- **The relationship between costs and quality varies by facility characteristics such as size and staffing, and the strategies which allow some facilities to provide quality care at a low cost are understudied given the prevalence of secondary data analyses in this literature.**
- **RN staffing and nursing case mix that favors licensed nurses is expensive and increases costs, but may be essential to improve quality.**

- **The relationship between structure, processes and outcome measures is likely not as strong as the current quality measurement system assumes, and various QI's are impacted differently by reimbursement changes. Generally studies supported spending on staffing and process measures.**

2. Influence of Value Based Purchasing programs on care quality outcomes

- **Providers respond variably to VBP incentives, and transparency/clarity regarding quality measurement is necessary to improve provider decision making.**
- **Perverse incentives exist in the system that may de-incentivize top facilities from improving quality.**
- **A single VBP threshold and weighting system for a state is possibly less effective than a more individualized, consultant style system that rewards facilities for addressing particular areas of quality concern.**
- **Overall, VBP systems improve quality in a less dramatic fashion than was anticipated when the programs began.**

3. NH report card and MDS quality measures

- **Clarity, simplicity and transparency regarding quality measurement is needed to increase resident and family engagement with the report card for decision making.**
- **There is evidence that consumer driven weighting and individualized composite measures are feasible and valid approaches to measuring quality.**
- **Public reporting of quality may result in disparities of nursing home self-selection. Those with high resources tend to cluster in facilities with high quality.**
- **Despite some concerns about accuracy of self-report measures, current MDS measures are generally well correlated with outcomes, stable and sensitive. However, some measures are considerably better at differentiating between high and low quality facilities than others, and QIs can perhaps be grouped into composite measures for simplification.**

Summary of study findings:

1. NH quality, reimbursement, and costs (5 secondary analyses of state data; 2 secondary analyses of Veterans' Administration (VA) data; 1 secondary analysis of Swiss data; 5 secondary data analyses of national MDS data; 1 mixed methods study)

Burgess et al. (2018) found the relationship between quality and costs within VA nursing homes varied by size and structure of the facility. Small facilities that improved clinical quality indicators had higher costs, while large facilities that improved had lower costs. No relationship was noted between costs and measures of resident centered care. Carey et al (2018) found that within VA nursing homes higher quality predicted higher costs, and lower quality predicted lower costs. The study contradicts others that found poor care outcomes such as falls, pressure ulcers and other inefficiencies led to higher costs. A study from within Swiss nursing homes found poor QI performance, specifically on pain and wt. loss, was related to higher costs, contributing to the evidence for an inverse relationship between costs and quality (DiGeorgio et al., 2016). Examining data from Missouri nursing homes, Hicks et al. (2004) found resident days accounted for the most variation in cost, indicating that provision of basic care, regardless of quality, impacts cost. Declining ADL's and pressure ulcers accelerated costs, demonstrating an inverse relationship similar to other studies. Mukamel and Spector (2000), examining trends in New York state data, noted a U-shaped relationship between quality and costs, with some high quality facilities having very low costs. Using Missouri data, Rantz et al. (2004) noted

higher costs in low quality facilities. Weech-Maldonado et al. (2006) found the relationship between cost and quality was not linear and differed based upon the quality outcome examined.

In many identified studies staffing appeared to be a relevant factor in the relationships between quality, reimbursement and costs. In an examination of 494 Texas nursing homes, Anderson et al. (1998) found no significant differences in spending allocation patterns between facilities with the best/ worst average outcomes. However, facilities with the highest improvement in resident outcomes had the highest costs and highest level of RN staffing. A more recent study in Ohio (Bowblis & Applebaum, 2017) found changes in state Medicaid reimbursement resulted in corresponding staffing changes, although quality indicators were not significantly affected. Authors proposed that something unmeasured at the micro level was perhaps occurring that drove the decision to spend on staffing despite the challenge of moving quantitative measures. A retrospective panel study of California nursing homes demonstrated mixed results in regards to costs, quality and staffing. Dulal (2017) found costs were inversely related to quality (lower costs, higher QI's), unrelated to inspection data, and higher staffing was related to cost inefficiency as defined by the study. Higher quality nursing homes had low costs, primarily due to fewer poor outcomes. Staffing was related to higher costs but not necessarily higher quality. Similarly, Grabowski (2001) found that among a national sample of nursing homes higher Medicaid reimbursement was related to increased nurse staffing but not an increase in quality. In a subsequent study, Grabowski et al. (2004) found higher reimbursement to be related to higher quality, although authors noted that the mechanism for the relationship was unclear. Weech-Maldonado et al. (2019), examining a national sample, found that higher RN staffing was related to high quality but lower financial performance, concluding that RN staffing may be needed, but at a cost.

2. Influence of Value Based Purchasing programs on care quality outcomes (1 survey of administrators; 1 retrospective analysis of CMS data; 2 retrospective analysis of multiple government sources)

In a multi-state of evaluation of the impact of VBP implementation on quality and costs, Grabowski et al. (2017) concluded that VBP had little impact on quality or costs, and that payments should be large enough to influence change and not simply reward already strong facilities. Adequate reimbursement to incentivize change emerged from a survey of 2,426 nursing home administrators from within 8 states with VBP policies and 8 states with no VBP policy. The survey found that administrators felt that quality is costly, and that VBP does not cover the cost. Respondents also questioned transparency of program administration and relevancy of measures to actual quality of care (Castle et al., 2014).

Werner et al. (2013) compared nursing home quality before and after VBP implementation in VBP and non-VBP states. Compared to non-VBP states, clinical quality measures improved, staffing was unchanged, and deficiencies increased, concluding that the impact of VBP was variable and inconsistent. Werner et al. (2016) investigated the impact of performance thresholds in pay for performance programs on nursing home response/ behavior. They measured nursing home performance in 6 states before and after threshold based VBP programs and found that most improvement was seen in the worst nursing homes, while the best nursing homes declined in quality. One study of hospitals (Das et al., 2016) was included in this summary because of its direct examination of VBP outcomes when the program emphasizes costs related to quality. Das et al. (2016) found adding an emphasis on costs/ spending in VBP for hospitals resulted in payments for efficiency that maintained quality, but also in payments to low quality hospitals that did not invest in improving care. Authors concluded that minimum quality thresholds are needed as not to reward providers for cost efficiency that does not maintain or improve quality.

3. NH report card and MDS quality measures (1 interviews with state program administrators; 3 commentary; 3 retrospective MDS analyses; 2 correlation between interview and MDS assessments; 3 secondary analyses of multiple government sources; 1 evaluation study; 2 survey; 1 secondary analysis of state data; 1 mixed method; 1 interviews with families)

Castle & Ferguson (2010), postulate that measurement of nursing home quality is highly intertwined with government regulation, and has evolved from minimum quality standards to a definition of quality aimed at reaching highest level of care. Current measures focus upon structure, process and outcomes which has both positive and negative influences on quality measurement. Risk adjustment, while necessary, also brings in limitations. The current search identified evidence which correlated the current quality ratings system to better actual resident outcomes. Cornell et al. (2019) found discharge to a higher star rated facility led to significantly lower mortality, fewer days in the nursing home, fewer hospital readmissions, and more days at home or with home health care during the first six months post facility admission. Results also indicated that within facility improvement results in improvement in resident outcomes. Rantz et al. (2004) investigated the ability of MDS-derived quality indicators to differentiate between high and low quality facilities in Missouri. They found that 10 of the QI's appeared to be sensitive to differentiating between facilities with poor and good quality outcomes and in general the MDS measures appeared stable. In a subsequent study Rantz et al. (2004) coupled observations of care with secondary data and found that consistency in basic care such as ambulation and nutrition were noted in facilities with good quality. Also, smaller facilities had better outcomes, and quality facilities had stable leadership and a team approach. Despite the evidence in support of MDS measures, Shanghavi et al (2019) found that 57% of resident falls with an acute care visit were reported on MDS, and facilities were less likely to report for non-white residents and in facilities with high proportion of non-white residents, as well as higher reporting rate for long stay than for short stay residents.

Several studies examined the influence of quality rating systems on nursing home selection. Konetzka (2014) found a correlation between financial ability and residence in a 5-star rated facility, with Medicaid eligible residents more likely to live in a 5-star home if they already lived there and the facility improved, as opposed to moving there, concluding that the 5-star policy inadvertently drove those with more choice to higher quality homes. Shapira (2016) conducted interviews with family members of newly admitted residents and found that when made aware of the report card people liked it, but more clarity is needed for the public to understand the methodology surrounding selection, measurement and weighting of quality scores. Similarly, Weimer et al. (2019) surveyed a sample of 4,310 residents to test the feasibility of using a consumer driven weighting approach instead of an expert determined weighting approach for the quality report card. They found staffing and inspection results to be the most consistent priorities of residents, with wide variation in the other QIs.

Drummond et al. (2015) matched interviewer assessments with MDS assessments and found strong correlation between the two assessments that remained stable even with MDS data collected 41 days from the interview assessment, providing additional evidence regarding the validity of MDS based quality measures. Mukamel et al. 2016 also examined use of MDS-derived measures for end of life care and found quantitatively valid measures, with the limitation that key aspects of patient choice are missing from the measure. Xu et al. (2016) conducted a factor analysis and concluded that summary measures could be created to adequately capture 4 dimensions of care quality. Kutschar et al. (2019) found item response to be stable in assessment among residents with mild cognitive impairment, but moderate cognitive impairment was negatively related to resident response in assessment. Pamalee et al. (2009) conducted an online survey of nursing home leaders and found that

ratings of the utility of MDS data were generally high, however qualitative findings suggested concerns around data accuracy, team functioning, timeliness of assessments, and validity of the MDS tool itself.

Interviews with administrators from 6 state value-based reimbursement programs revealed that measurement of quality varied between the 6 states, with some common measures. The most common approach to financial award based upon quality was a daily add on to the Medicaid rate (Arling et al, 2009). Konetzka et al. (2018) found evidence that facilities improved what was emphasized by the quality rating system, with higher weight placed on clinical measures correlated to improvements in those areas, but low weighting being correlated with decline in those areas, and skilled staffing increasing when weight placed on staffing. Both high and low quality homes were influenced by incentive program weighting of quality measures. Arling et al. (2009) highlighted the need to move beyond 'one size fits all' quality measurement, an idea that was validated by Mukamel et al.'s (2016) evaluation of a demonstration project comparing personalized selection of measures, weighting and subsequent rankings with the 'one size fits all' model. They found that personalized measures differed enough between individuals and from CMS that such a model should be considered for nursing home selection.

An expert commentary (Miller & Mor, 2008) noted the need for better, more specific data and more facility-specific and quality improvement focused regulation that is consistent between states, regions, and districts within states. In an earlier commentary, Mor (2005) noted that a risk of composite measures is that some facilities perform well on one, poorly on another, and when the average is taken the facilities appear equal; important differences are missed. Using data to motivate quality improvement is especially challenging, as even under controlled conditions QI's are hard to move. Mor (2005) suggests that context effects such as leadership may be the true driver of change.
References

Anderson, R. A., Hsieh, P. C., & Su, H. F. (1998). Resource allocation and resident outcomes in nursing homes: Comparisons between the best and worst. Research in Nursing & Health, 21(4), 297-313.

Arling, G., Job, C., & Cooke, V. (2009). Medicaid nursing home pay for performance: where do we stand?. The Gerontologist, 49(5), 587-595.

Bowblis, J. R., & Applebaum, R. (2017). How does Medicaid reimbursement impact nursing home quality? The effects of small anticipatory changes. Health services research, 52(5), 1729-1748.

Burgess Jr, J. F., Shwartz, M., Stolzmann, K., & Sullivan, J. L. (2018). The relationship between costs and quality in veterans health administration community living centers: an analysis using longitudinal data. Health Services Research, 53(5), 3881-3897.

Carey, K., Zhao, S., Snow, A. L., & Hartmann, C. W. (2018). The relationship between nursing home quality and costs: Evidence from the VA. PloS one, 13(9).

Castle, N. G., & Ferguson, J. C. (2010). What is nursing home quality and how is it measured?. The Gerontologist, 50(4), 426-442.

Castle, N. G., Engberg, J., Ferguson-Rome, J. C., & Sonon, K. (2014). Nursing Home Administrators' Opinions of Pay for Performance. Journal of aging & social policy, 26(3), 229-248.

Cornell, P. Y., Grabowski, D. C., Norton, E. C., & Rahman, M. (2019). Do report cards predict future quality? The case of skilled nursing facilities. Journal of health economics, 66, 208-221.

Das, A., Norton, E. C., Miller, D. C., Ryan, A. M., Birkmeyer, J. D., & Chen, L. M. (2016). Adding a spending metric to Medicare's value-based purchasing program rewarded low-quality hospitals. Health Affairs, 35(5), 898-906.

Drummond, L. S., Slaughter, S. E., Jones, C. A., & Wagg, A. S. (2015, September). Affirming the value of the Resident Assessment Instrument: Minimum Data Set Version 2.0 for nursing home decision-making and quality improvement. In Healthcare (Vol. 3, No. 3, pp. 659-665). Multidisciplinary Digital Publishing Institute.

Dulal, R. (2017). Cost efficiency of nursing homes: do five-star quality ratings matter?. Health care management science, 20(3), 316-325.

Di Giorgio, L., Filippini, M., & Masiero, G. (2016). Is higher nursing home quality more costly?. The European Journal of Health Economics, 17(8), 1011-1026.

Grabowski, D. C. (2001). Does an increase in the Medicaid reimbursement rate improve nursing home quality?. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 56(2), S84-S93.

Grabowski, D. C., Angelelli, J. J., & Mor, V. (2004). Medicaid payment and risk-adjusted nursing home quality measures. Health Affairs, 23(5), 243-252.

Grabowski, D. C., Stevenson, D. G., Caudry, D. J., O'Malley, A. J., Green, L. H., Doherty, J. A., & Frank, R. G. (2017). The impact of nursing home pay-for-performance on quality and medicare spending: results from the nursing home value-based purchasing demonstration. Health services research, 52(4), 1387-1408.

Hicks, L. L., Rantz, M. J., Petroski, G. F., & Mukamel, D. B. (2004). Nursing home costs and quality of care outcomes. Nursing Economics, 22(4), 178-192.

Tamara Konetzka, R., Grabowski, D. C., Perraillon, M. C., & Werner, R. M. (2015). Nursing home 5-star rating system exacerbates disparities in quality, by payer source. Health affairs, 34(5), 819-827.

Konetzka, R. T., Skira, M. M., & Werner, R. M. (2018). Incentive design and quality improvements: Evidence from state Medicaid nursing home pay-for-performance programs. American journal of health economics, 4(1), 105-130.

Kutschar, P., Weichbold, M., & Osterbrink, J. (2019). Effects of age and cognitive function on data quality of standardized surveys in nursing home populations. BMC geriatrics, 19(1), 244.

Miller, E. A., & Mor, V. (2008). Balancing regulatory controls and incentives: Toward smarter and more transparent oversight in long-term care. Journal of Health Politics, Policy and Law, 33(2), 249-279.

Mor, V. (2005). Improving the quality of long-term care with better information. The Milbank Quarterly, 83(3), 333-364.

Mukamel, D. B., & Spector, W. D. (2000). Nursing home costs and risk-adjusted outcome measures of quality. Medical Care, 38(1), 78-89.

Mukamel, D. B., Amin, A., Weimer, D. L., Sharit, J., Ladd, H., & Sorkin, D. H. (2016). When patients customize nursing home ratings, choices and rankings differ from the government's version. Health Affairs, 35(4), 714-719.

Mukamel, D. B., Ladd, H., Caprio, T., & Temkin-Greener, H. (2016). Prototype end-of-life quality measures based on MDS 3 data. Medical care, 54(11), 1024-1032.

Parmelee, P. A., Bowen, S. E., Ross, A., Brown, H., & Huff, J. (2009). "Sometimes people don't fit in boxes": attitudes toward the minimum data set among clinical leadership in VA nursing homes. Journal of the American Medical Directors Association, 10(2), 98-106.

Rantz, M. J., Hicks, L., Petroski, G. F., Madsen, R. W., Mehr, D. R., Conn, V., ... & Maas, M. (2004). Stability and sensitivity of nursing home quality indicators. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 59(1), M79-M82.

Rantz, M. J., Hicks, L., Grando, V., Petroski, G. F., Madsen, R. W., Mehr, D. R., ... & Bostick, J. (2004). Nursing home quality, cost, staffing, and staff mix. The Gerontologist, 44(1), 24-38.

Sanghavi, P., Pan, S., & Caudry, D. (2019). Assessment of nursing home reporting of major injury falls for quality measurement on nursing home compare. Health Services Research.

Schapira, M. M., Shea, J. A., Duey, K. A., Kleiman, C., & Werner, R. M. (2016). The nursing home compare report card: perceptions of residents and caregivers regarding quality ratings and nursing home choice. Health services research, 51, 1212-1228.

Weech-Maldonado, R., Shea, D., & Mor, V. (2006). The relationship between quality of care and costs in nursing homes. American Journal of Medical Quality, 21(1), 40-48.

Weech-Maldonado, R., Pradhan, R., Dayama, N., Lord, J., & Gupta, S. (2019). Nursing home quality and financial performance: is there a business case for quality?. INQUIRY: The Journal of Health Care Organization, Provision, and Financing, 56, 0046958018825191.

Weech-Maldonado, R., Lord, J., Pradhan, R., Davlyatov, G., Dayama, N., Gupta, S., & Hearld, L. (2019). High Medicaid Nursing Homes: Organizational and Market Factors Associated With Financial Performance. INQUIRY: The Journal of Health Care Organization, Provision, and Financing, 56, 0046958018825061.

Weimer, D. L., Saliba, D., Ladd, H., Shi, Y., & Mukamel, D. B. (2019). Using contingent valuation to develop consumer-based weights for health quality report cards. Health services research, 54(4), 947-956.

Werner, R. M., Skira, M., & Konetzka, R. T. (2016). An evaluation of performance thresholds in nursing home pay-for-performance. Health services research, 51(6), 2282-2304.

Werner, R. M., Konetzka, R. T., & Polsky, D. (2013). The effect of pay-for-performance in nursing homes: evidence from state Medicaid programs. Health services research, 48(4), 1393-1414.

Xu, D., Kane, R. L., Shippee, T., & Lewis, T. M. (2016). Identifying consistent and coherent dimensions of nursing home quality: Exploratory factor analysis of quality indicators. Journal of the American Geriatrics Society, 64(12), e259-e264

.

Citation	Study Objective	Study Design	Sample and Data	Outcome of Interest	Findings	Limitations	Implications
Anderson, R. A., Hsieh, P. C., & Su, H. F. (1998). Resource allocation and resident outcomes in nursing homes: Comparisons between the best and worst. *Research in Nursing & Health*, *21*(4), 297-313.	To examine and compare resource allocation patterns within the best and worst performing nursing homes in Texas; Do high quality nursing homes allocate financial resources differently than low quality nursing homes?	Secondary data analysis from State database	494 nursing homes divided into 5 comparison groups based upon resident outcome measures	Resource allocation differences between nursing homes with high and low quality resident outcomes	There was no significant differences in spending allocation patterns between facilities with the best/ worst average outcomes. If RN spending is controlled for, quality outcomes did not vary by high/ low cost facilities. Facilities with the highest improvement in resident outcomes had the highest costs and highest level of RN staffing.	Single state, all measures not available	RN's contribute to improving care, and are a high cost investment in improving quality outcomes. When RN staffing is controlled for in the analysis, the relationship between cost and quality improvement decreases. If investing and increasing costs, nursing staff is likely to most impact quality. Impact of spending/ allocation on static measures such as averages is more difficult to determine.

Arling, G., Job, C., & Cooke, V. (2009). Medicaid nursing home pay for performance: where do we stand?. *The Gerontologist*, *49*(5), 587-595.	To provide a snapshot of current pay for performance programs in nursing homes and provide recommendations based upon the experiences of 6 states	Structured interviews with administrators of the 6 state nursing home pay for performance programs operating in 2007	See study design	Structure and administration of current (2007) nursing home pay for performance	Measurement of quality varied between the 6 states, with some common measures. The most common approach to financial award is a daily add on to the Medicaid rate.	Findings were highly interpretive	Evidence-based measurement, clear predictable paths to achieve reward, stakeholder input, and state support for an overall quality plan that goes beyond financial incentives were among the recommendations
Bowblis, J. R., & Applebaum, R. (2017). How does Medicaid reimbursement impact nursing home quality? The effects of small anticipatory changes. *Health services research*, *52*(5), 1729-1748.	To examine how anticipated changes in Medicaid rates impacts nursing home spending and resident outcomes; how changes in state reimbursement impact quality	Retrospective secondary analysis of government databases	All Ohio nursing homes	Spending allocation, costs, and resident outcomes	Changes to reimbursement as a result of state policy varied among Ohio nursing homes, allowing for comparison based upon reimbursement change. Changes in reimbursement led to	Retrospective and reliant on existing measures; cannot account for the lag in quality changes	Reimbursement changes resulted in corresponding staffing changes. However, quality indicators were not significantly affected. Something unmeasured at the micro level is perhaps occurring that

					correspondin g changes (up or down) in staffing. However, no significant changes in resident outcomes were noted.		**drives the decision to spend on staffing despite challenge of moving quantitative measures**
Burgess Jr, J. F., Shwartz, M., Stolzmann, K., & Sullivan, J. L. (2018). The relationship between costs and quality in veterans health administration community living centers: an analysis using longitudinal data. *Health Services Research*, *53*(5), 3881-3897.	**To determine the relationship between costs and quality**	**Retrospectiv e secondary data analysis**	**130 VA nursing homes over a 3 year period**	**Clinical quality indicators from the MDS, measures of resident centered care**	**Small facilities that improved clinical quality indicators had higher costs, large facilities that improved had lower costs. No relationship was noted between costs and measures of resident centered care.**	**No information regarding allocation decisions**	**The relationship between costs and quality varies by size and structure. High quality may require high costs, or in other settings high quality is the result of efficient lower cost processes. Mixed methods work is needed.**
Carey, K., Zhao, S., Snow, A. L., & Hartmann, C. W.	**To examine the relationship between costs**	**Retrospectiv e secondary data**	**135 VA nursing homes over**	**Costs**	**Higher quality predicted**	**Aggregate only facility**	**The study contradicts others that**

(2018). The relationship between nursing home quality and costs: Evidence from the VA. *PloS one*, *13*(9).	and resident outcomes in VA nursing homes	analysis, MDS outcomes	a 2 year period		higher costs, lower quality predicted lower cost	level VA specific data	found poor care outcomes such as falls, pressure ulcers and other inefficiencies leads to higher costs and supports a basic economic argument that good care is expensive. It did not however find that high costs were being allocated to nurse staffing.
Castle, N. G., & Ferguson, J. C. (2010). What is nursing home quality and how is it measured?. *The Gerontologist*, *50*(4), 426-442.	Overview and commentary of nursing home quality measurement	Framed the discussion using Donabedian's structure, process and outcome framework	See study design	none	Measurement of nursing home quality is highly intertwined with government regulation, and has evolved from minimum quality standards to a definition of quality	Findings are highly interpretive	The relationship between structure, process and outcomes is likely not as strong as the current quality measurement structure assumes. Risk adjustment accounts for uneven 'playing

					aimed at reaching highest level of care. Current measures focus upon structure, process and outcomes which is both positive and negative. Risk adjustment, while necessary, also brings in limitations.		fields' between facilities but may obfuscate some real quality differences. Linearity is also assumed and likely not reflective of true quality differences. The link between measuring quality and improving quality remains uncertain.
Castle, N. G., Engberg, J., Ferguson-Rome, J. C., & Sonon, K. (2014). Nursing Home Administrators' Opinions of Pay for Performance. *Journal of aging & social policy*, *26*(3), 229-248.	To examine nursing home administrators perceptions of pay for performance incentive structures, program administration, and quality measurement/ impact	Mail survey of nursing home administrators in 8 states that had implemented VBR and 8 randomly selected states that had not	Surveys from 2,426 respondents almost evenly divided between VBR states and non VBR states	Respondent opinions of pay for performance	Overall perceptions were very low. Respondents felt payments should be higher, more frequent, and more reflective of the costs to improve quality. Measurement was viewed	Likert scale survey that left the rating scale up to the interpretation of respondents.	Administrators felt that quality is costly, and that VBR does not cover the cost. Also questioned transparency of program administration and relevancy of measures to actual quality of care.

					as not transparent and not related to actual quality. Opinions within states with VBR were significantly lower than in states without VBR. Paper provides table of perceptions regarding which indicators should be included or excluded from quality composites.		
Cornell, P. Y., Grabowski, D. C., Norton, E. C., & Rahman, M. (2019). Do report cards predict future quality? The case of skilled nursing	To determine the relationship between quality star ratings and resident outcomes, and to contribute to	Retrospective analysis of MDS and Medicare claims data, community data, other	Claims, geographic and MDS data from 1,278,456 Medicare beneficiaries discharged	Resident outcome disposition: rehospitalization, death, hospice, home	Discharge to a higher star SNF led to significantly lower mortality, fewer days in the nursing	Big data analysis makes a number of assumptions and does not account for factors	Star quality ratings are reflective of quality in terms of resident trajectory/ disposition, and should be a

facilities. *Journal of health economics*, *66*, 208-221.	the quality literature an analysis that accounts for the contribution of resident selection bias of high/ low quality facilities to the relationship between quality and outcomes	secondary data	from 4,332 acute care hospitals to 15,166 SNFs.	with home health	home, fewer hospital readmissions, and more days at home or with home health care during the first six months post SNF admission. Results also show that within facility improvement results in improvement in resident outcomes	that are unmeasured, such as discharge planner influence on choice	part of resident's decision making processes.
Das, A., Norton, E. C., Miller, D. C., Ryan, A. M., Birkmeyer, J. D., & Chen, L. M. (2016). Adding a spending metric to Medicare's value-based purchasing program rewarded low-quality hospitals. *Health*	To determine that impact of VBR policy that emphases costs/ low spending over quality measures on the distribution of rewards	Retrospective analysis of CMS databases and American Hospital Association data	CMS data on 2,679 hospitals eligible in 2014-2015	Financial incentive received by hospital	Adding an emphasis on costs/ spending in VBR for hospitals resulted in payments for efficiency, but also payments to low quality hospitals that	Secondary data analysis that cannot capture unmeasured variance	Minimum quality thresholds are needed as not to reward providers for cost efficiency that does not results in quality.

Affairs, *35*(5), 898-906.					did not invest in care		
Drummond, L. S., Slaughter, S. E., Jones, C. A., & Wagg, A. S. (2015, September). Affirming the value of the Resident Assessment Instrument: Minimum Data Set Version 2.0 for nursing home decision-making and quality improvement. In *Healthcare* (Vol. 3, No. 3, pp. 659-665). Multidisciplinary Digital Publishing Institute.	To compare interview completed functional assessments with the functional assessment recorded on the MDS for consistency	Correlational analysis	362 paired interviewer assessments and MDS assessments from 130 nursing home residents	Stability of MDS measure when compared to more comprehensive interviewer assessment	MDS assessment was correlated to interviewer assessment and remains stable even with MDS collected 41 days from interview	Data collection for both measures subject to interviewer bias	Adds confidence to MDS function measures and QI's
Dulal, R. (2017). Cost efficiency of nursing homes: do five-star quality ratings matter?. *Health care management science*, *20*(3), 316-325.	To investigate what factors influence nursing home costs, and how quality influences costs	Retrospective quantitative analysis	Panel survey of California nursing homes from 2009-13 with n ranging from 761-919. Data included quality	Nursing home costs	Costs were inversely related to quality (lower costs, higher QI's), unrelated to inspection data, and higher	Single state data, secular influences on cost unmeasured	High costs do not necessarily mean high quality, and investment in process change instead of simply higher staff can

			measures, inspection data and staffing levels		staffing was related to cost inefficiency as defined by the study. Higher quality nursing homes had low costs, primarily due to fewer poor outcomes. Staffing was related to higher costs but not necessarily higher quality		improve cost efficiency.
Di Giorgio, L., Filippini, M., & Masiero, G. (2016). Is higher nursing home quality more costly?. *The European Journal of Health Economics*, *17*(8), 1011-1026.	To determine the relationship between quality and costs in Swiss nursing homes	Retrospective quantitative analysis	Data from 45 Swiss nursing homes between 2006-10, including QI's (the IV) and costs	Nursing home costs	Poor QI performance, specifically on pain and wt. loss, was related to higher costs; process measure performance was not related to costs	Results may not be transferable to US healthcare system	Reimbursement systems should account for a relationship between quality and costs that varies based upon quality measure, and that high costs do not mean high quality

Grabowski, D. C. (2001). Does an increase in the Medicaid reimbursement rate improve nursing home quality?. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *56*(2), S84-S93.	To examine the relationship between changes in Medicaid reimbursement and nursing home quality	Retrospective data analysis of linked government data sets	Facility level data from a national sample of >15K facilities	Nursing home quality measures	Increased Medicaid rate improved the level of professional staffing, but not other quality measures; increased rates decreased deficiencies in tight economic markets but not overall	Secondary analysis of facility data leaves much unmeasured	Higher reimbursement may encourage better staffing but not necessarily better care.
Grabowski, D. C., Angelelli, J. J., & Mor, V. (2004). Medicaid payment and risk-adjusted nursing home quality measures. *Health Affairs*, *23*(5), 243-252.	To examine the relationship between Medicaid reimbursement rate and quality	Retrospective analysis of linked government data sets	Facility level data from a national sample of >15K facilities	Nursing home quality measures	Higher payment was related to lower pressure ulcer and restraint rates. Authors note that the 2 measures are not well correlated so it may indicate better quality	Many unmeasured variables, limited quality measure assessment to 3	Higher reimbursement may result in higher quality, though the mechanism is unclear

					across the spectrum		
Grabowski, D. C., Stevenson, D. G., Caudry, D. J., O'Malley, A. J., Green, L. H., Doherty, J. A., & Frank, R. G. (2017). The impact of nursing home pay-for-performance on quality and Medicare spending: results from the nursing home value-based purchasing demonstration. *Health services research*, *52*(4), 1387-1408.	To evaluate the impact of VBR on quality and Medicare spending	Retrospective analysis of quantitative data from baseline measures to measures from within a 3 year VBR demonstration project; qualitative staff interviews	Facility data from New York facilities randomized into the demonstration and matched demonstration facilities in WI and AZ	Nursing home quality measures; Medicare spending rates	No changes in Medicare spending or quality were noted within the NY facilities; facilities in WI and AZ had Medicare savings for part of the time period. Interviews noted few changes were made within facilities due to demonstration, and respondents perceived that already existing quality was simply being rewarded, instead of encouraging	Differences between state contexts may not have been fully controlled for in the analysis	VBR demonstration had little impact on quality or costs. Payments should be large enough to influence change an not just reward already strong facilities.

					new quality efforts		
Hicks, L. L., Rantz, M. J., Petroski, G. F., & Mukamel, D. B. (2004). Nursing home costs and quality of care outcomes. *Nursing Economics*, *22*(4), 178-192.	To examine the relationship between variable costs and 4 QI's: ADL decline, pressure ulcers, psychotropic drug use, weight loss	Secondary analysis of linked MDS and Medicaid cost reports	474 Missouri nursing homes	Variable nursing home costs	Resident days accounted for the most variation in cost, indicating that provision of basic care, regardless of quality, impacts cost. Declining ADL's and pressure ulcers accelerated costs.	Single state, not indepth enough to know what factors are increasing costs	Poorer care quality defined by resident decline results in higher variable cost of providing adequate care. However, most cost contributes to providing basic adequate care, regardless of variation in quality.
Tamara Konetzka, R., Grabowski, D. C., Perraillon, M. C., & Werner, R. M. (2015). Nursing home 5-star rating system exacerbates disparities in quality, by payer source. *Health affairs*, *34*(5), 819-827.	To determine if public reporting of quality measures resulted in more non-dual eligibles selecting high quality homes and more dual eligibles	Retrospective quantitative design of linked government data sets	Linked MDS, Nursing home compare, Medicare claims for US nursing homes	Dual eligibles residing in high and low quality nursing homes	The gap between duals and non duals in high quality homes grew over time since reporting began, and duals were more likely to live in a high	Multiple assumptions are made about nursing home selection in the interpretation of findings	Supply of homes and location of high quality homes matters, 5 star policy inadvertently drove those with more choice to higher quality homes, raising Medicaid rates

	residing in low quality homes				quality home because the 5 star rating improved, as opposed to moving there		to be more equitable with private rates is a possible solution
Konetzka, R. T., Skira, M. M., & Werner, R. M. (2018). Incentive design and quality improvements: Evidence from state Medicaid nursing home pay-for-performance programs. *American journal of health economics*, *4*(1), 105-130.	To examine how design of state pay for performance incentive programs influences nursing home quality improvements	Retrospective analysis of government data sets	Linked MDS, state quality reporting data, and data program data for all US nursing homes, including 3,472 (20%) in VBR states	Facility level quality, health inspection and staffing levels over time	Higher weight placed on clinical measures causes improvements in those areas, but low weight actually causes decline in those areas; minimum deficiency thresholds are more effective than weighting deficiencies on the incentive structure; skilled staffing increases when weight	Analysis did not provide information on the processes that may be influencing these relationships, though reasons were hypothesized	Weights influence quality behavior of facilities, and programs should perhaps weight most heavily what is needed by a particular facility as opposed to applying the same incentive structure to all

					placed on staffing; both high and low quality homes were influenced by incentive programs		
Kutschar, P., Weichbold, M., & Osterbrink, J. (2019). Effects of age and cognitive function on data quality of standardized surveys in nursing home populations. *BMC geriatrics*, *19*(1), 244.	To determine if resident characteristics, particularly cognitive impairment, influence the quality of survey data among nursing home residents	Analyzed survey data collected from pre/post intervention to determine influences on non-response	659 residents within 13 German nursing homes	Item non-response	Interview duration and gender had no effect, age had a mild effect, and level of cognitive impairment had a significant effect with a significant difference between mild and moderate impairment	Only non-response, not validity of response, was measured	Even with face to face survey/ interview methods, moderate cognitive impairment can negatively influence survey data quality
Miller, E. A., & Mor, V. (2008). Balancing regulatory controls and incentives: Toward smarter and more transparent oversight in long-	To provide expert commentary on the current regulatory process and potential areas	Commentary	None, past research	None	Regulatory is crucial, but current practices suffer from limited data, a 'one size	Commentary only (but a really good one)	We need better, more specific data and more facility-specific and quality improvement

term care. *Journal of Health Politics, Policy and Law*, *33*(2), 249-279.	of improvement				fits all' mentality, and a punitive relationship between providers and states. There is also great inconsistency between states, and political influence from the nursing home lobby varies between states to influence the system. An improved model would use facility data to advise facilities how to improve an reward that improvement, much like a consultant		focused regulation. Regulation should be more consistent between states, regions, and districts within states.

Mor, V. (2005). Improving the quality of long-term care with better information. *The Milbank Quarterly*, *83*(3), 333-364.	**To describe the use of data to measure nursing home quality**	**Essay/ commentary**	**None**	**None**	**Data/ information can incentivize quality by impacting consumer choices, reward structures, and/or punishment. Essay usefully describes types of quality information such as individual vs. aggregate, process vs. outcome, establishing quality benchmarks, and risk adjustment for comparisons**	**Commentary only (but a really good one)**	**A risk of composite measures is that some facilities perform well on one, poorly on another, and when the average is taken the facilities appear equal; important differences are missed. Using data to motivate quality improvement is especially challenging, as even under controlled conditions QI's are hard to move. Context effects such as leadership may be the true driver of change.**

Mukamel, D. B., & Spector, W. D. (2000). Nursing home costs and risk-adjusted outcome measures of quality. *Medical Care*, *38*(1), 78-89.	To understand the relationship between quality and costs in nursing homes and to test the hypothesis that higher quality is related to lower costs	Secondary data analysis of New York State database	525 nursing homes within NY state	1. risk adjusted pressure ulcers, ADL decline, and mortality 2. variable costs	A non-linear U shaped relationship between quality and costs suggesting some high quality facilities have low costs	Only 3 quality measures and limited definition of costs for analytic purposes	Financial restraints does not always mean (or need to mean) low quality; strategies which result in low cost high quality care need further identification
Mukamel, D. B., Amin, A., Weimer, D. L., Sharit, J., Ladd, H., & Sorkin, D. H. (2016). When patients customize nursing home ratings, choices and rankings differ from the government's version. *Health Affairs*, *35*(4), 714-719.	To compare data with 146 residents who used the individualized nursing home compare plus composite measure with the CMS composite measure	Demonstration project comparing personalized selection of measures, weighting and subsequent rankings with the 'one size fits all' model	146 patients and families (42 were patients) who were discharged from hospital to the nursing home	Difference between measures, weighting, rankings	Almost all users (97%+) selected PT and nurse staffing in their measure; high variability among other measures; <15% chose restraints or catheters; substantial disagreement between CMS and CMSplus	May not be a feasible approach to VBP	Personalized measures differed enough between individuals and from CMS that such a model should be considered for nursing home selection
Mukamel, D. B., Ladd, H., Caprio, T., & Temkin-Greener,	To develop and test end of life quality	Secondary data analysis	39,590 nursing home	Death in the hospital, number of	End of life QMs had variation	Misses key measures of patient	The MDS could provide some valid data for

H. (2016). Prototype end-of-life quality measures based on MDS 3 data. *Medical care*, *54*(11), 1024-1032.	**measures from MDS data**	**of NY state database**	**decedents in 626 facilities in NY state**	**hospitalizations, pain, and depression during the last 90 days before death**	**across facilities similar to that observed for other QMs The pain and depression QMs were significantly better among nursing homes ranked as 4 and 5 stars compared with those ranked as 1 and 2 stars for most dimensions. The hospitalizations QMs were significantly better among nursing homes**	**choice, advanced directives, and emotional care**	**end of life measures**

					with a higher staffing rating.		
Parmelee, P. A., Bowen, S. E., Ross, A., Brown, H., & Huff, J. (2009). "Sometimes people don't fit in boxes": attitudes toward the minimum data set among clinical leadership in VA nursing homes. *Journal of the American Medical Directors Association*, *10*(2), 98-106.	To describe attitudes toward the MDS among nursing home unit leadership in the VA	Online survey with some open ended items	289 directors of nursing, medical directors, MDS coordinators, nurse managers	Perception of MDS: accuracy, usefulness for QI, reasons for inaccuracy or non-use	Ratings were generally high, however qualitative findings suggested concerns around data accuracy, team functioning, timeliness of assessments, and validity of the MDS tool itself. MD's were least favorable, as were very large and very small facilities	Only VA system, no objective measures	Respondents appeared to appreciate MDS data but noted multiple weaknesses in its utility
Rantz, M. J., Hicks, L., Petroski, G. F., Madsen, R. W., Mehr, D. R., Conn, V., ... & Maas, M.	To determine the ability of nursing home QI's to detect differences in	Retrospective analysis of secondary	92 randomly selected Missouri	23 quality indicators; stability of performance over time and	10 of the QI's appeared to be sensitive to differentiatin	Single state, outcome measurement may not be truly	The 10 identified QI's may be best to use when

(2004). Stability and sensitivity of nursing home quality indicators. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *59*(1), M79-M82.	quality between nursing homes and describe the quality of the nursing home	government data sets	nursing homes	sensitivity to quality outcomes/ use in classifying facilities	g between facilities with poor and good quality outcomes and in general the MDS measures appear stable	reflective of quality	classifying facilities
Rantz, M. J., Hicks, L., Grando, V., Petroski, G. F., Madsen, R. W., Mehr, D. R., ... & Bostick, J. (2004). Nursing home quality, cost, staffing, and staff mix. *The Gerontologist*, *44*(1), 24-38.	To describe the processes of care, organizational attributes, cost of care, staffing level, and staff mix in a sample of Missouri homes with good, average, and poor resident outcomes	Mixed methods: retrospective analysis of large secondary government data sets; observations of care processes	92 randomly selected Missouri nursing homes divided into 3 comparison groups based upon quality rating	Observed care processes; structural attributes of facilities; total cost per resident day	Observed consistency in basic care such as ambulation and nutrition were noted in facilities with good quality; smaller facilities had better outcomes; quality facilities had stable leadership and a team approach; costs were higher in	Single state study	Quality appears to depend more on leadership and team processes then spending/ costs

					poor quality facilities and staffing/ staff mix did not vary between groups		
Sanghavi, P., Pan, S., & Caudry, D. (2019). Assessment of nursing home reporting of major injury falls for quality measurement on nursing home compare. *Health Services Research*.	To assess the accuracy of the MDS reports of major injury falls and determine facility characteristics that may be associated with under reporting of falls	Linked claims and MDS data, multi-level modeling	150,828 major fall reports within a national sample (100%) of nursing home residents' with Medicare claims	Correlation between acute care claims and MDS fall report	57% of acute care claim falls were reported on MDS; less likely to report for non-white residents and in facilities with high proportion of non-white residents; reporting higher for long stay than short stay residents	The use of claims data may miss some falls, or may overestimate that number of falls that occurred in the facility	The MDS falls measure may be inaccurate
Schapira, M. M., Shea, J. A., Duey, K. A., Kleiman, C., & Werner, R. M. (2016). The nursing home compare report card: perceptions of	To evaluate the perceived usefulness of the report card to residents and families	Primary data collection, structured interviews	Convenience sample of 35 residents (6) or families (29) newly admitted to the nursing home in the	Perceptions of star ratings, comparisons, and use of the report card for decision making	Positive perception of quality information overall but confusion over how the quality was	Convenience sample in a single geographic area	When made aware of the report card people like it, but more clarity is needed for the public to

residents and caregivers regarding quality ratings and nursing home choice. *Health services research*, *51*, 1212-1228.			Philadelphia area		actually measured and the relationship between domain specific and overall quality score		understand the methodology
Weech-Maldonado, R., Shea, D., & Mor, V. (2006). The relationship between quality of care and costs in nursing homes. *American Journal of Medical Quality*, *21*(1), 40-48.	To evaluate the impact of providing quality care on nursing home costs	Secondary data analysis of government data sets	749 nursing homes in 5 states	Total patient care costs per facility	Neither QI was linear to costs. Pressure ulcers was an inverted U with costs lower for higher quality after a threshold; mood decline was a flat curve for low quality with increasing costs for higher quality	Only 2 QI's were examined, and cost measure was not specific enough to fully explore implications	The relationship between cost and quality is not linear and differs based upon the quality outcome examined
Weech-Maldonado, R., Pradhan, R., Dayama, N., Lord, J., & Gupta, S. (2019). Nursing home quality and financial	To determine the relationship between nursing home quality and	Secondary analysis of government data sets	All free standing non-government nursing	Total operating margin per facility	Apart from staffing (structure), nursing homes that have better	Secondary data sets may miss crucial	An investment in staffing is expensive but may be

performance: is there a business case for quality?. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, *56*, 0046958018825191.	financial performance		homes in the US		processes and outcomes have better financial performance	processes of care	necessary; poor quality is costly
Weech-Maldonado, R., Lord, J., Pradhan, R., Davlyatov, G., Dayama, N., Gupta, S., & Hearld, L. (2019). High Medicaid Nursing Homes: Organizational and Market Factors Associated With Financial Performance. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, *56*, 0046958018825061.	To examine factors that correlate with better financial performance among high Medicaid nursing homes	Retrospective data analysis of large secondary data sets	Approximately 1108 high Medicaid facilities per study year	Nursing home operating and total margin	Higher financial performing facilities have more beds, are for profit, in low competition markets, and higher occupancy; RN staffing related to lower financial performance	Secondary data, unable to determine relationships beyond correlations	Staffing is expensive but may be needed for high quality; having slack resources such as occupancy and little competition may allow for low resource innovation
Weimer, D. L., Saliba, D., Ladd, H., Shi, Y., & Mukamel, D. B. (2019). Using	To test the feasibility of using a consumer	Web survey asking "willingness to trade"	4310 nursing home residents or	Calculated "willingness to trade" to weight QI's	Respondent's choices appear economically	Unusual method that has a number of	The trade off method may be useful to inject consumer

contingent valuation to develop consumer-based weights for health quality report cards. *Health services research*, *54*(4), 947-956.	driven weighting approach instead of an expert determined weighting approach for the quality report card	visit/ travel time to facility for quality in specific measures	recent residents	based upon consumer preferences	rationale based but vary considerably between QI and respondent characteristics. The most largest weighting was staffing and inspections.	assumptions regarding the perceived trade off value of travel time	priorities into QI measures
Werner, R. M., Skira, M., & Konetzka, R. T. (2016). An evaluation of performance thresholds in nursing home pay-for-performance. *Health services research*, *51*(6), 2282-2304.	To investigate the impact of performance thresholds in pay for performance programs on nursing home response/ behavior	Retrospective analysis of secondary government data sets	Nursing homes within 6 states implementing pay for performance, with one set used as subjects and second set of 3 for comparison	Performance before and after implementation of threshold based programs	The most improvement was seen in the worst nursing homes, while the best nursing homes declined in quality	Programs vary by state	There is the potential for perverse incentives in threshold based programs that may discourage high performing facilities from improving, but low performing facilities appear motivated by the program
Werner, R. M., Konetzka, R. T., & Polsky, D. (2013). The effect of pay-for-performance in	To test the impact of pay for performance program	Retrospective MDS and OSCAR analysis	Nursing homes in 8 states implementing pay for	Change in nursing home quality after policy	Compared to non P4P states, clinical quality	State programs varied in	Impact of P4P on quality was variable and inconsistent

nursing homes: evidence from state Medicaid programs. *Health services research*, *48*(4), 1393-1414.	**implementation on nursing home quality**		**performance, with the other 42 states as controls**	**implementation**	**measures improved, staffing was unchanged, and deficiencies increased**	**timing and composition**	
Xu, D., Kane, R. L., Shippee, T., & Lewis, T. M. (2016). Identifying consistent and coherent dimensions of nursing home quality: Exploratory factor analysis of quality indicators. *Journal of the American Geriatrics Society*, *64*(12), e259-e264.	**To determine if there are consistent dimensions of QI's that are stable at the resident and facility levels**	**Retrospective analysis of secondary government data sets**	**Residents admitted to 382 Minnesota nursing homes in 1 year period**	**Dimensions of QI's**	**4 dimensions were identified, and they remained consistent between the resident and facility level**	**Single state study**	**Summary measures can be created to capture care quality**

Chapter 4

Cost and Quality Trajectory Clustering

Zachary Hass, PhD

Executive Summary

Purpose

It has been hypothesized that greater spending on direct care should lead to higher quality care. However, the quantitative link between care cost and care quality has been difficult to establish at the macro level. The Value Based Reimbursement (VBR) policy change, effective January 1, 2016, incentivized greater spending on care related costs with the expectation that this would contribute to greater quality of care. An increase in spending on care related to VBR has been observed, but not a similar related rise in quality scores. The purpose of this analysis is to identify sub-groups of facilities in terms of care related cost and quality score trajectories. These sub-groups may indicate differing strategies or constraints that facilities use and face when responding to the VBR policy.

Methods

Latent Class Growth Analysis is used to simultaneously cluster facilities by their care related cost and quality score trajectories. Once an optimally fitting model is found, these clusters are then compared across a range of facility characteristics to better understand the sub-groups. Hospital attached facilities were removed from the clustering analysis and presented in the tables as a separate cluster.

Results

- **Three clusters were identified: Cluster 3 maintains relatively higher costs, began with relatively high quality and ended in the middle of the clusters for mean quality score. Cluster 1 and Cluster 2 tracked fairly closely with relatively lower costs, but Cluster 1 began and ended with relatively higher quality scores while Cluster 2 began with low quality, improved in the middle of the period, and declined in quality scores at the end of the period.**
- **Cluster 2 (relatively lowest quality scores) is characterized by a higher rate of CHOW, for-profit ownership, relatively higher administrative costs per resident day, relatively low revenue to long term lease ratios (for those facilities with long term leases), lower overall staff retention, scored relatively worse on all quality measures, and spent relatively less on group medical insurance per resident day.**
- **Cluster 1 (relatively low cost and highest quality) is notably similar to Cluster 2 in many facility and spending characteristics not otherwise noted, but has the best quality scores for hospitalization rate per 1000 resident days, quality indicator score, MDH inspection score, and overall quality score. Cluster 1 has relatively much more favorable total revenue to long term lease costs (for those facilities with long term leases) and spends the most on group medical insurance per resident day.**
- **Cluster 3 (relatively highest cost, middle quality) has a much higher average number of admissions, made up of mostly non-profit and government facilities, higher acuity and occupancy, lowest percentage of Medicaid days, almost entirely located in the Metro area, highest care related costs, best staff retention, best adjusted community discharge and hospitalization rates, and highest quality of life scores (marginally).**
- **Cost trajectories are more stable (smooth) than quality score trajectories.**

This section of the report describes the analysis that explores the relationship between quality and care related costs for subsets of facilities. Latent Class Growth Analysis was utilized to better understand the sub-population structure of Minnesota nursing facilities in terms of spending on direct care costs and quality scores. Data were derived from the 2013-2019 Facility Cost Reports and data from the Minnesota Quality Report Cards. Quadratic growth models were used and the number of clusters was set to a range from 2-5, and various fit criterion (eg. AIC/BIC) and estimation quality (eg. condition number) were assessed for the best fitting model. The model with three clusters was superior. Cost and quality were modeled jointly as the response and the algorithm iteratively assigned each facility to a cluster in order to maximize model fit of the parameters. All modeling results in this report were fit on non-hospital attached facilities, although the hospital attached facilities are displayed as a comparison cluster.

Initial Clusters

Table 1: Mean Growth Parameters from the Three Clusters

Cluster	Quality Intercept	Quality Slope	Quality Quadratic	Cost Intercept	Cost Slope	Cost Quadratic	N
1	76.59	0.75	-0.12	99.52	3.24	0.66	163
2	70.85	3.93	-0.75	98.44	7.90	-0.01	87
3	77.09	0.82	-0.24	126.09	4.57	0.26	40

Table 1 displays the estimated mean growth parameters for the three clusters. Cluster size ranges from 40-163 facilities. The table also gives the intercept (starting quality or care related cost in year 2013), slope (linear growth component), and quadratic term (quadratic growth component). The quadratic model is used as the linear model proved to be inadequate as several facilities show exponential, logarithmic, and parabolic growth or decline patterns. For example, Cluster 2 has a quality intercept of 70.85, slope of 3.93, and quadratic term of -0.75. The intercept indicates that the cluster mean at the beginning of the period (2013) was a quality score of 70.85. The combination of a slope of 3.93 and quadratic term of -0.75 indicates a pattern of initial growth with a peak in the middle then a decline in the latter half of the time period (inverted "U" shape). This translates to an increase of 3.18 points in 2014, 1.68 points in 2015, 0.18 points in 2016, a decrease (-1.32 points) in 2017, (-2.82) in 2018, and (-4.32) in 2019. This trend and other cluster trends is represented graphically in the appendix (red line in spaghetti plots).

Mean Patterns in Care Related Cost and Quality Scores for Largest Clusters

Figure 1 and Figure 2 display the mean trajectories for care related costs and overall quality score for the 3 clusters. In both figures, the means are calculated as the average across the facilities within a cluster for each year and the trend line is a simple connecting of the dots. Figure 2 indicates that there is a general upward trend in costs with Cluster 3 maintaining relatively higher costs in the group, cluster 2 and 3 track more closely, beginning and ending nearer to each other than in the middle of

the period. Mean quality scores for the group tended to be higher than 70 until 2018-19 when cluster 2 drops into the 60s. Despite having the highest costs Cluster 3 had a small decline in quality scores towards the end of the period (2017-19) leaving them as the middle performing cluster. Despite the lowest costs cluster 1 finished with the highest quality scores of the three groups.

Appendix Spaghetti Plots

Spaghetti plots by cluster are included in the appendix (Figure A3) to illustrate the estimated trajectories and level of variability among the facilities. Two spaghetti plots are included for each cluster, quality score and care related cost trajectory. Each facilities values are represented by an individual line, the solid red line gives the estimated cluster trajectory from the growth model. For example, cluster three shows an inverted parabola for quality (upside down U) and linear cost growth.

Cluster Characteristics

Table 2 through Table 6 describe the 3 clusters and hospital attached facilities using facility characteristics (Table 2), spending patterns (Table 3), staffing patterns (Table 4), quality metrics (Table 5), and employee compensation metrics (Table 6). These tables can help characterize or tell the story of each cluster. Cluster 2 is of primary interest as Figure 1 indicates that although care related costs of Cluster 2 track fairly closely to Cluster 1, quality scores are much lower and seem to have decreased relative to the other two groups over the last three years (2017-2019). Cluster 2 will be compared for characteristics relative to the other facilities.

Table 2 shows that Cluster 2 facility characteristics are similar in many respects to Cluster 1, but stands out as having the highest rate of ownership change (37% vs 22% for cluster 1 / 13% for cluster 3), for profit facilities (46% vs 38/20%), and having increasing occupancy over the last year (+1.5% vs -0.8/-0.2%). Both Cluster 2 and Cluster 1 lag behind Cluster 3 (relatively highest cost and middle quality) number of admissions (408 vs 141/164) and occupancy rate (90% vs 84%). Cluster 3 is notably 73% non-profit facilities, has the lowest percentage of Medicaid days (51% vs 56/59%), highest rate of Other RUG Paid days (40% of non-MA days vs 28/30%), and is primarily located in the Metro area (95% vs 66/63%).

Table 3 shows that the mean spending patterns of Cluster 1 and 2 track fairly closely while Cluster 3 tends to spend more on care related costs per standardized day ($162 vs $142/144). Cluster 3's total spending in the aggregate are notably higher on Central Office and Other Administrative Costs ($889K vs $435K/$495K) and Net Administrative Costs less Insurance/Working Capital/Bad Debt ($1326K vs $716K/$774K). However, when adjusting these costs per standardized resident day, cluster 2 was highest for both categories ($21.28 vs $19.41/$20.99 and $33.28 vs $31.96/$31.30 respectively). Importantly, 30% of Cluster 2 facilities report long term lease costs with an average ratio of total revenue per long term lease cost of $115, comparable to 17% of Cluster one with $162,695 of total revenue to long term lease cost and 23% of Cluster 3 with $2,004 of total revenue to long term lease cost.

Table 4 displays staffing patterns across the clusters. Retention is measures as the percentage of employees from beginning of the cost year (October 1) remaining employed by the facility at the end of the cost year (September 30). Cluster 2 is the lowest in overall retention (64% vs 67/72%) retention of nurse administrators (71% vs 79/73%), RNs (66% vs 71/74%), LPNs (70% vs 73/74%), social workers (59% vs 77/81%), activities staff (67% vs 75/80%), and other direct care staff (25% vs 32/62%). Compensated direct care hours per resident day are noticeably higher for Cluster 3 in RNs (0.88 vs

0.63/0.65) in total for direct care (5.25 vs 4.82/4.84), and for social workers (0.2 vs 0.15/0.14), but don't differ greatly between the three groups in the other categories.

Table 5 presents the mean quality scores and measures for the clusters. Cluster 2 has the lowest scores for those measures that were used in the clustering (expected as a function of the model) and for those measures not used in the clustering (not necessarily expected). Cluster two was lowest on overall quality score (66.9 vs 77.1/73.4) as well as each sub-component, quality indicators (30.5 vs 36.7/34), quality of life (31.7 vs 32.5/32.7) and the Minnesota Department of Health score (4.7 vs 7.9/6.7). Cluster two was also lowest on the measures not used for trajectory clustering such as adjusted community discharge within 30 days of admission (31% vs 34/39%), between 31-90 days (31% vs 33/35%), adjusted re-hospitalization rates within 30 days of admission (12.5% vs 12.3/11.9%), and unadjusted hospitalization rate per 1000 resident days (1.75 vs 1.53/1.55).

Table 6 displays compensation related measures across the clusters such as salary, insurance and benefits, and union presence. Cluster 2 spends the least on group medical insurance in the aggregate ($267 vs $298K/$555K) and on the per resident day basis ($12.02 vs $13.58/$13.16). Salary per resident day for Cluster 2 tends to be above Cluster 1 and below Cluster 3 with the exception of CNAs ($40.81 vs $41.34/$43.64), direct care trainer ($3.37 vs $3.42/$4.61), mental health worker ($3.43 vs $3.58/$5.31), and social worker were the order is reversed ($4.92 vs $5.37/$4.90).

Summary

This report represents the best fitting model using all non-hospital attached facilities with cost report and quality data from 2013-2019. It is possible that removing certain outlying facilities from the analysis, that tend to form small or individual clusters when larger numbers of clusters are used in the modeling, may lead to a well-fitting model with greater resolution (larger number of clusters/sub-populations of facilities). The current analysis highlights three clusters. Cluster 1 and Cluster 2 have similar cost trajectories, while Cluster 3 had substantively higher mean costs. Cluster 1 maintained the highest mean quality, while Cluster 3 settled into the middle at the end of the period, and Cluster 2 initially closed the gap and then fell away to a greater gap at the end then they began with. Although Cluster 2 has comparable costs to Cluster 1, Cluster 2 appears to struggle more with retention which may be in part due the larger proportion of facility change in ownership and slightly lower benefit compensation such as health insurance. Notably cluster 2 also has much less favorable long term lease positions relative to the other clusters. Understanding the mechanisms underlying these problems will be imperative to remedying the quality gap.

Figure 1: Mean Facility Quality Score and Cost Trajectory by Assigned Cluster Membership

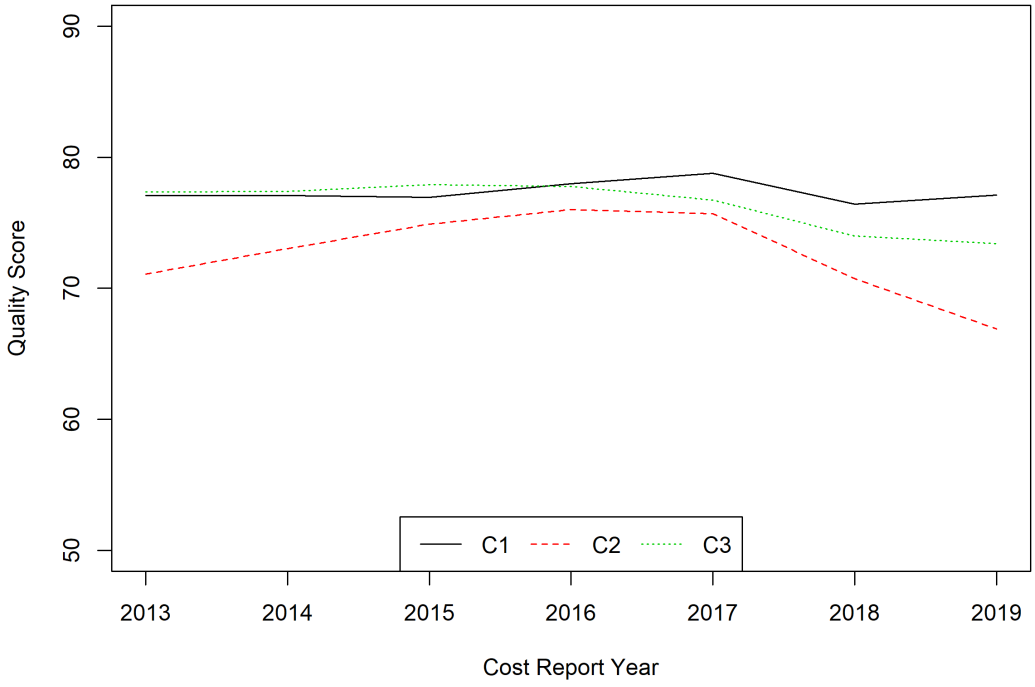


Figure 2: Mean Care Related Cost Trajectories by Assigned Cluster Membership

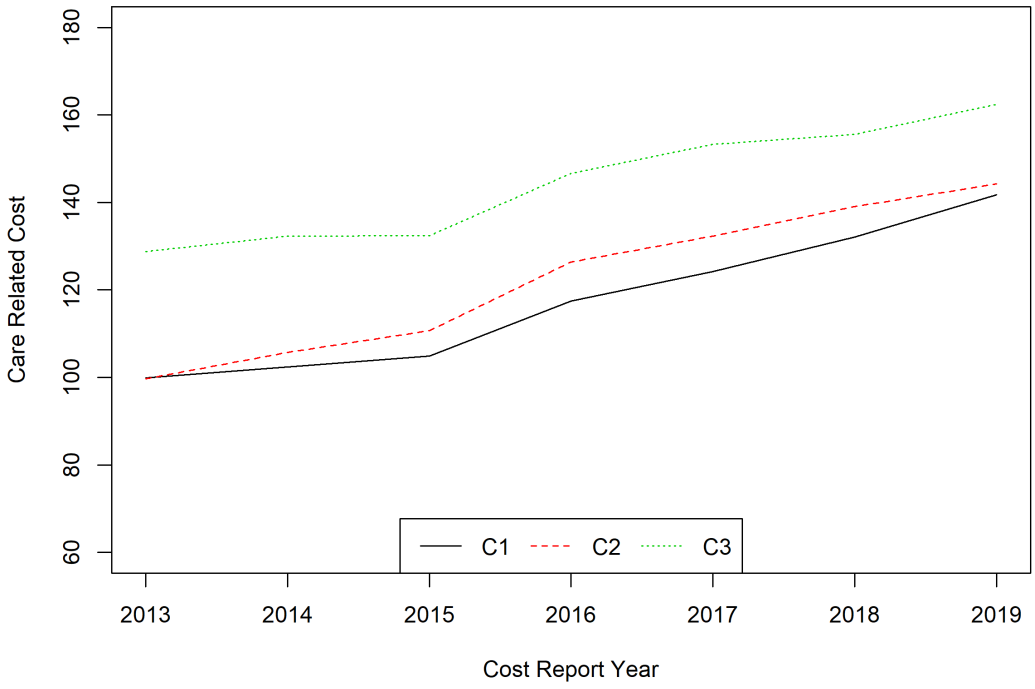


Table 2: Facility Characteristics by Cost/Quality Trajectory Cluster

Cluster	C 1	C 2	C 3	Hospital Attached
Facilities	163	87	40	43
Annual Admissions	140.8	163.8	408.2	74.8
Ownership Change	22%	37%	13%	7%
Ownership: For Profit	38%	46%	20%	2%
Ownership: Non Profit	56%	47%	73%	74%
Ownership: Government	6%	7%	8%	23%
Hospital Attached*	0%	0%	0%	100%
Acuity	1.01	1.02	1.04	0.92
Latest Annual Occupancy %	84%	84%	90%	86%
Latest Annual Occupancy Change	-0.8%	1.5%	-0.2%	-0.8%
Resident Days	22,016	22,565	40,841	18,599
Medicaid Paid Resident Days	12,373	13,390	20,937	11,638
Medicaid Days / Resident Days	56%	59%	51%	63%
Non-Medicaid Paid Resident Days	9,644	9,175	19,904	6,961
Medicare RUG Paid Resident Days	1,747	1,882	4,235	855
Medicare RUG / Non-MA Days	18%	21%	21%	12%
Other RUG Paid Residents Days	2,672	2,784	7,995	1,725
Other RUG / Non-MA Days	28%	30%	40%	25%
Private Pay RUG Paid Resident Days	5,224	4,509	7,674	4,382
Private Pay RUG / Non-MA Days	54%	49%	39%	63%
Total Beds	71.2	73.1	124.9	58.5
Metropolitan RUCA	66%	63%	95%	44%
Micropolitan RUCA	15%	15%	3%	19%
Small Town RUCA	13%	16%	0%	14%
Rural RUCA	6%	6%	3%	23%
Medicare LTC Basket Wage Index	0.99	0.99	1.10	0.93

Cluster	C 1	C 2	C 3	Hospital Attached
Standardized Days	22,417	23,259	42,372	17,561
Minnesota Only Facility (single state)	71%	66%	60%	93%
Chain Facility	63%	67%	65%	78%

***Hospital attached facilities grouped into their own cluster manually for comparison purposes.**

Table 3: Facility Spending Patterns by Cost/Quality Trajectory Cluster

Cluster	C1	C2	C3	Hospital Attached
Facilities	163	87	40	43
Direct Care Cost per Standardized Day	116	120	132	146
Other Care Related Cost per Standardized Day	25.5	24.4	30.6	31.1
Total Care Cost per Standardized Day	142	144	162	177
Other Operating Cost per Resident Day	77.0	78.1	78.9	101.1
Dietary Cost per Resident Day	15.5	15.3	17.4	23.8
Laundry Cost per Resident Day	3.9	3.5	3.7	4.8
Housekeeping Cost per Resident Day	7.4	7.1	8.3	6.6
Physical Plant Cost per Resident Day	15.6	15.8	15.6	9.8
Administrative Cost per Resident Day	34.6	36.4	34.0	56.1
Administrative Management Fees per Resident Day	9.7	9.8	12.7	0.5
Central Office and General Admin Other (K)	435	495	889	62
Central Office and General Admin Other PSRD	19.41	21.28	20.99	3.54
Net Admin less Insurance, Working Capital, Bad Debt (K)	716	774	1,326	1,025
Net Admin less Insurance, Working Capital, Bad Debt PSRD	31.96	33.28	31.30	58.36
Direct Care Cost per Resident Day	118	122	137	134
Other Care Related Cost per Resident Day	25.5	24.7	30.5	28.5
Total Care Related Cost per Resident Day	143	147	168	162
Long Term Lease Cost (K)	53.0	119.9	43.6	6.1
Total Revenue per Long Term Lease Cost	162,695	115	2,004	271
Facilities with Long Term Lease	17%	30%	23%	7%

Table 4: Facility Staffing Patterns by Cost/Quality Trajectory Cluster

Cluster	C1	C2	C3	Hospital Attached
Facilities	163	87	40	43
Staffing Pool Use Percentage	0%	0%	0%	0%
Nursing Pool RN Direct Car Hours (K)	15	25	23	14
Nursing Pool LPN Direct Care Hours (K)	25	27	23	29
Nursing Pool CNA Direct Care Hours (K)	43	38	47	57
Nursing Pool TMA Direct Care Hours	33	9	6	77
Overall Staff Retention	67%	64%	72%	74%
Retention: Nurse Admin	79%	71%	73%	78%
Retention: RN	71%	66%	74%	70%
Retention: LPN	73%	70%	74%	80%
Retention: CNA	60%	60%	66%	68%
Retention: TMA	50%	53%	52%	51%
Retention: MHW	1%	0%	2%	0%
Retention: Social Work	77%	59%	81%	83%
Retention: Activities	75%	67%	80%	87%
Retention: ODC	32%	25%	62%	31%
LPN Compensated DC Hours per Resident Day	0.63	0.67	0.68	0.65
RN Compensated DC Hours per Resident Day	0.63	0.65	0.88	0.72
CNA Compensated DC Hours per Resident Day	2.39	2.33	2.40	2.58
Licensed Compensated DC Hours per Resident Day	1.26	1.32	1.55	1.37
Total Compensated DC Hours per Resident Day	4.82	4.84	5.25	5.19
Activities Staff Compensated DC Hours per Resident Day	0.32	0.29	0.26	0.39
Mental Health Workers Compensated DC Hours per Resident Day	0.00	0.00	0.03	0.00
Nursing Administration Compensated DC Hours per Resident Day	0.32	0.34	0.38	0.30
Social Workers Compensated DC Hours per Resident Day	0.15	0.14	0.20	0.12

Cluster	C1	C2	C3	Hospital Attached
Trained Medication Aides Compensated DC Hours per Resident Day	0.30	0.32	0.31	0.28
Other Direct Care Staff Compensated DC Hours per Resident Day	0.03	0.02	0.05	0.07

Table 5: Facility Quality Patterns by Cost/Quality Trajectory Cluster

Clusters	C1	C2	C3	Hospital Attached
Facilities	163	87	40	43
Adjusted Community Discharge Rate (3-30 Days)	34%	31%	39%	38%
Adjusted Community Discharge Rate (31-90 Days)	33%	31%	35%	32%
Adjusted Re-hospitalization Rate (3-30 Days)	12.3%	12.5%	11.9%	12.2%
Hospitalization Rate per 1000 Resident Days	1.53	1.75	1.55	1.14
Overall Quality Score	77.11	66.89	73.42	73.70
Quality Indicators (Scaled out of 50)	36.69	30.51	34.00	33.15
Minnesota Department of Health Quality Score (out of 10)	7.88	4.66	6.69	7.73
Quality of Life Quality Score (out of 40)	32.53	31.72	32.74	32.82

Table 6: Employee Compensation Patterns by Cost/Quality Trajectory Cluster

Clusters	C1	C2	C3	Hospital Attached
Facilities	163	87	40	43
Total Employer Health Insurance Expenditures (K)	302	259	555	625
Group Medical Insurance (K)	298	267	555	453
Nursing Admin Union	1%	2%	5%	2%
RN Union	6%	13%	15%	21%
LPN Union	17%	22%	25%	35%
CNA Union	21%	33%	38%	30%
Trained Medical Ast Union	19%	28%	23%	12%
Med Records Union	6%	7%	8%	5%
Mental Health Union	1%	0%	0%	0%
Social Workers Union	1%	1%	5%	2%
Activities Staff Union	17%	24%	20%	21%
Other DC Staff Union	4%	10%	8%	7%
Pharmacy Union	1%	2%	0%	0%
DC Staff Trainers Union	1%	7%	0%	0%
Direct Care Salary per Resident Day	96.62	100.47	115.09	106.87
Nursing Admin Salary per Resident Day	11.09	11.84	13.01	12.13
RN Salary per Resident Day	20.62	21.80	30.16	25.66
LPN Salary per Resident Day	17.43	19.17	20.91	17.55
CNA Salary per Resident Day	41.34	40.81	43.64	46.18
TMA Salary per Resident Day	5.47	6.04	6.11	5.25
Activities Salary per Resident Day	0.67	0.80	1.26	0.10
Direct Care Trainer Salary per Resident Day	3.42	3.37	4.61	3.28
Medical Records Salary per Resident Day	0.02	-	0.59	-
Mental Health Worker Salary per Resident Day	3.58	3.43	5.31	3.25
Social Worker Salary per Resident Day	5.37	4.92	4.90	6.90

Clusters	C1	C2	C3	Hospital Attached
Scholarship Cost per resident Day	1.12	0.87	1.15	0.78
Group Medical Insurance Cost per Resident Day	13.58	12.02	13.16	23.80

Appendix –

Spaghetti Plots of Care Related Costs and Quality Scores

Figure A1 and Figure A2 display the individual facility observed mean patterns in quality scores (Figure A1) and care related cost (Figure A2). Each line on the plot gives a facilities observed quality score and care related costs, the trend line connects the observations. Common colors indicate those facilities were clustered together. The primary takeaways from these plots are that there is more heterogeneity in quality scores than costs (i.e. costs show a stronger pattern of stratification by color), more volatility in the quality scores (i.e. the pattern over time is less clear), and more outlying observations in the quality scores.

Figure A1: Spaghetti Plot of Facility Quality Score Trajectory by Assigned Cluster Membership

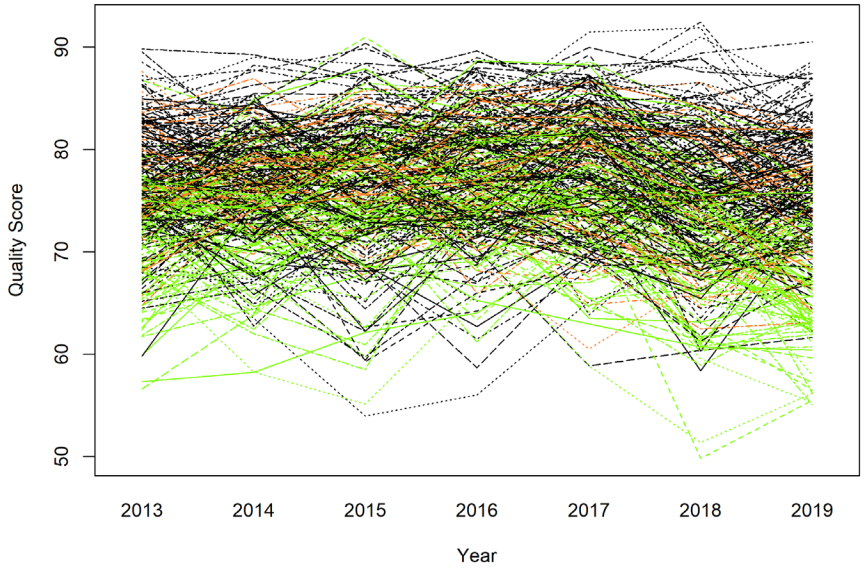


Figure A2: Spaghetti Plot of Facility Care Related Cost Trajectory by Assigned Cluster Membership

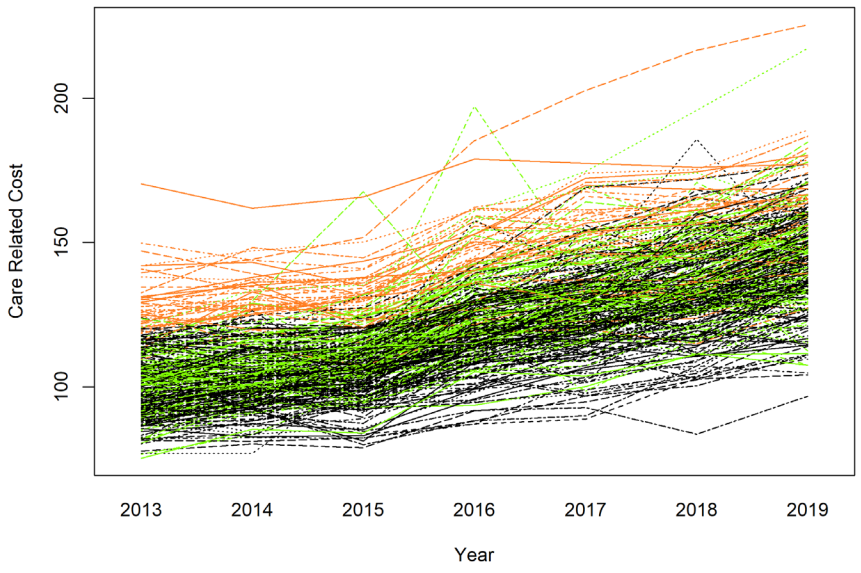
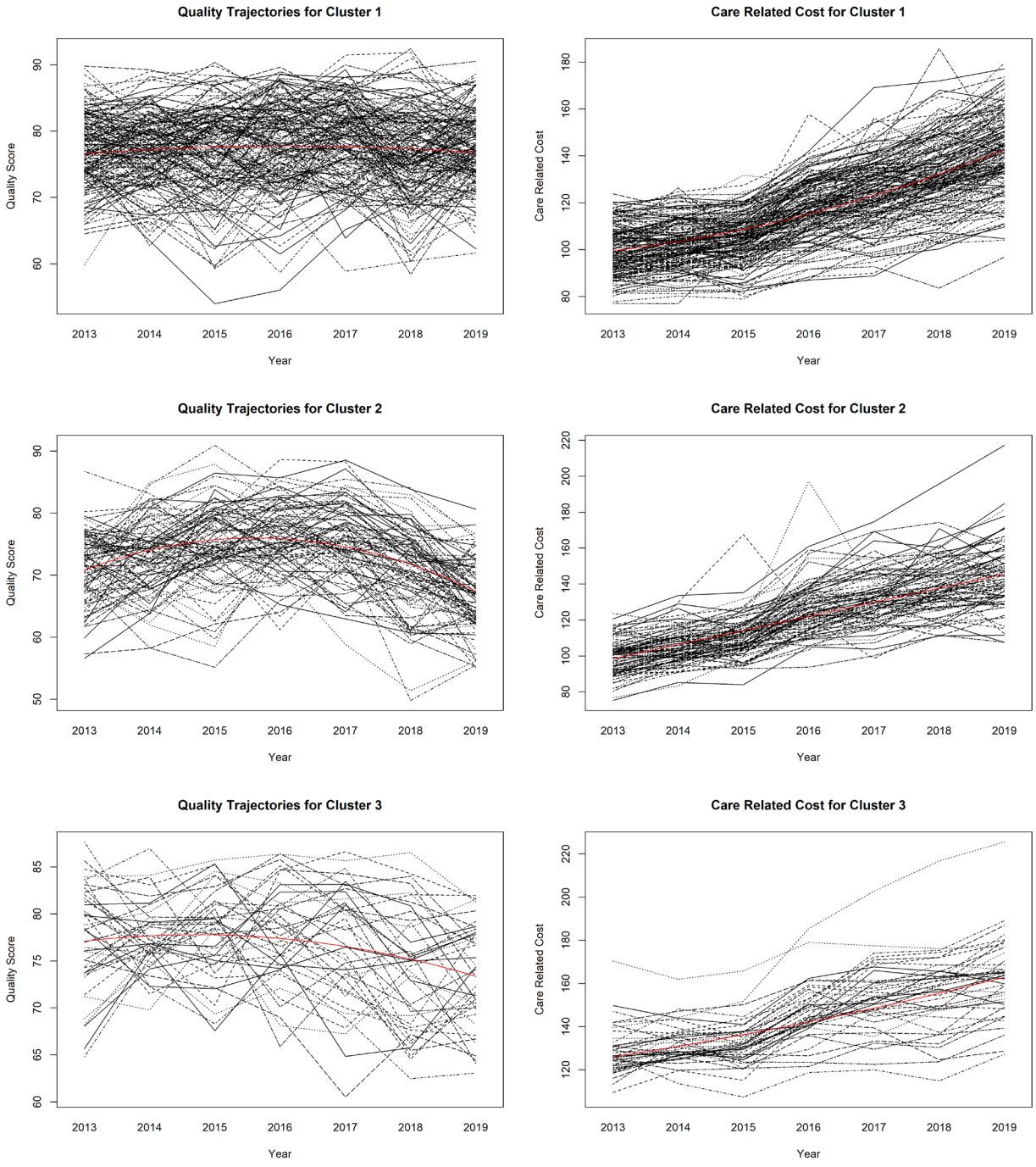


Figure A3: Spaghetti Plots of Cost and Quality for Each Cluster (Red Line Gives Modeled Trend)



Chapter 5

A Brief Update on Care Related Costs and Quality Scores by Ownership Change Status

Zachary Hass, PhD

Executive Summary

Purpose

The Value Based Reimbursement (VBR) legislation, effective January 1st 2016, coincided with a substantial uptick in change of ownership (CHOW) of nursing facilities in the state, often times from out of state buyers and/or buyers with convoluted ownership structures. It appears that VBR has made Minnesota facilities appear more attractive to profit seeking and it is desirable to understand whether or not these purchasing organizations are committed to upholding and/or improving quality of care for Minnesota's nursing facility residents. The purpose of this section is to investigate whether the CHOW event is correlated with substantive changes in facility quality and/or costs.

Methods

Data are from Nursing Facility Cost Reports and Minnesota Quality Report Card data from the 2013-2019 cost years (partially unaudited cost data in 2019). Data on CHOW facilities and related information provided from internal Minnesota Department of Human Services' (DHS) internal database. Various cost and quality metrics are compared between CHOW and non-chow facilities using line plots (between groups) and cost and quality overtime for individual CHOW facilities are also visualized using line plots (within group). The year prior and year after the CHOW event are used to compare metrics for possible changes in cost or quality directly related to the CHOW (within group comparison).

Results

- **CHOW facilities perform worse relative to facilities with constant ownership on every quality related metric (quality scores and subcomponents, staffing, retention, community discharge, hospitalization).**
- **At least some of the gap between groups is due to a selection effect as future CHOWs tend to be performing worse on quality metrics in each year than current and past CHOWs.**
- **For those facilities that have accumulated data following the CHOW event, there is a visually discernable downwards trend in quality for a majority of facilities.**
- **CHOW facilities appear to have reduced spending on Laundry and dental benefits, while increasing spending more slowly on medical and scholarship benefits, and increasing administrative management fees at a higher rate.**
- **The gap in quality scores between CHOW and non-CHOW facilities increases after a CHOW event.**

Introduction

When comparing the Change in Ownership (CHOW) group to the Same Owner group, there are two basic mechanisms that can lead to differences between the groups over time. The first is that facilities within the group are improving/declining in a given metric at a different rate than the other group over time (eg. if the CHOW group is performing more poorly over time than the same owner group, the gap in mean scores for a metric will grow over time). The second mechanism that can create a gap between the two groups is if there is a selection bias in which sort of facilities go through a CHOW (eg. if poorer performing facilities are more likely to be sold, than there will be a gap between the groups over time as poorer performing facilities are added into the CHOW pool). We present two plots for each metric, one with a constant CHOW group (all facilities with a CHOW between 2014-2019 are in the CHOW group for all years) and a variable CHOW group (facilities enter on or after their sale cost year). Change due to facilities should be most prevalent in the constant CHOW group plot and can be calculated as a difference-in-difference metric ([2019 Same group mean – 2013 Same group mean] – [2019 CHOW group mean – 2013 CHOW group mean]). Change due to selection can be seen by comparing the Variable CHOW group plot to the constant CHOW group plot. How much of the gap between groups is due to facilities in the constant CHOW group who have not yet been sold can be calculated for a given year as:

$$\frac{(2014\ CHOW\ mean_{Variable} - 2014\ CHOW\ mean_{Constant}) - (2014\ Same\ mean_{variable} - 2014\ Same\ mean_{constant})}{(2014\ Same\ mean_{constnat} - 2014\ Chow\ mean_{constant})}$$

The ownership change group includes 78 facilities with an ownership change and continuous operation during the period, and not attached to a hospital (2014 cost year – 5 facilities, 2015 – 7, 2016 – 13, 2017 – 26, 2018 – 19, 2019-8). For the constant group comparison the same owner group increased care related costs per standardized day by $5.45 more than the CHOW group over the period (difference-in-difference metric). Most of this increase occurred in the 2019 cost year. The percentage of the gap in the constant group plot due to inclusion of future CHOWs (getting at the idea of selection) by year is: 2014 – 168%, 2015 – 32%, 2016 – 30%, 2017 – 16%, 2018 – 4% (and 2019 – 0%). In 2014, the percentage being greater than 100% indicates that the gap was reversed by the inclusion of future CHOWs (the black line is vertically higher in the variable CHOW group plot on the right), meaning that future CHOWs had much lower costs than the 2014 cohort of CHOWs. By 2017, future CHOWs had similar cost structure to the CHOW cohort (as future CHOWs made up 16% or less of the gap).

Figure 3: Care Related Cost Trajectories by Ownership Change Status

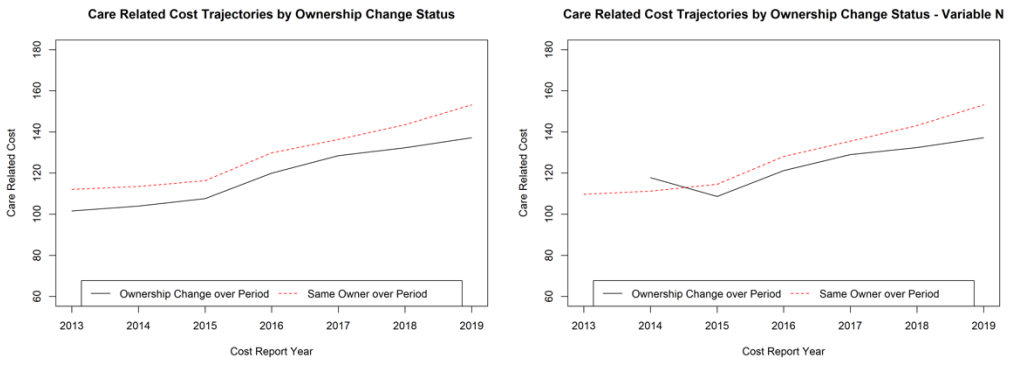
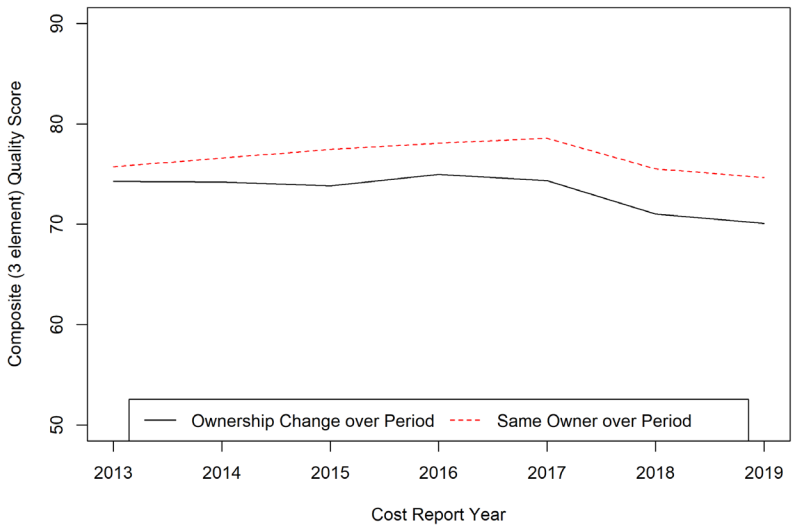


Figure 4: Quality Score Trajectories by Ownership Change Status (Constant Groups)



Despite the fairly close tracking of mean costs, the two groups began with similar mean quality scores and had a growing gap over time (diff-in-diff metric – the same owner group had a 3.11 quality point increase over the time period relative to the CHOW group). The annual percentage of the gap between the two groups due to including future CHOWs in the CHOW group: 2014 – 12%, 2015 – 53%, 2016 – (-11%), 2017 – 20%, 2018 – 22%. In 2016, future CHOWs (those sold in 2017-2019) were actually performing slightly better than the CHOW group (sold 2014-2016) leading to a reduction in the gap in the constant CHOW group plot (Figure 4).

Figure 5 Quality Score Trajectories by Ownership Change Status (Cumulative CHOW Group)

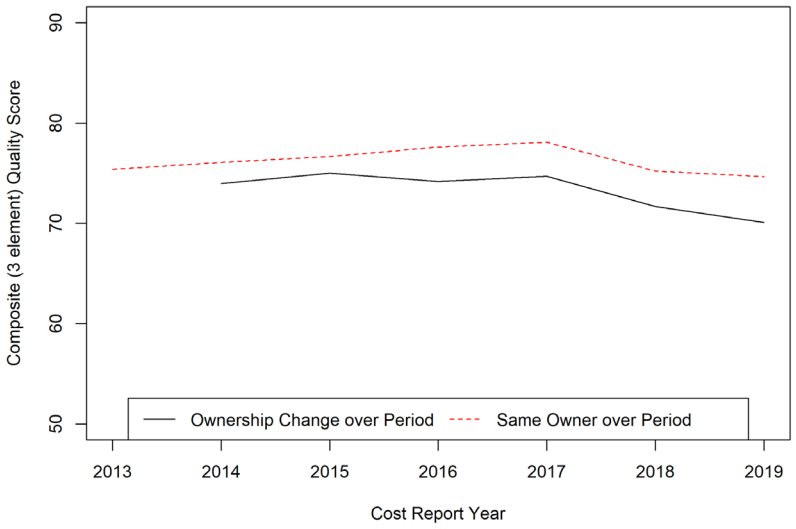
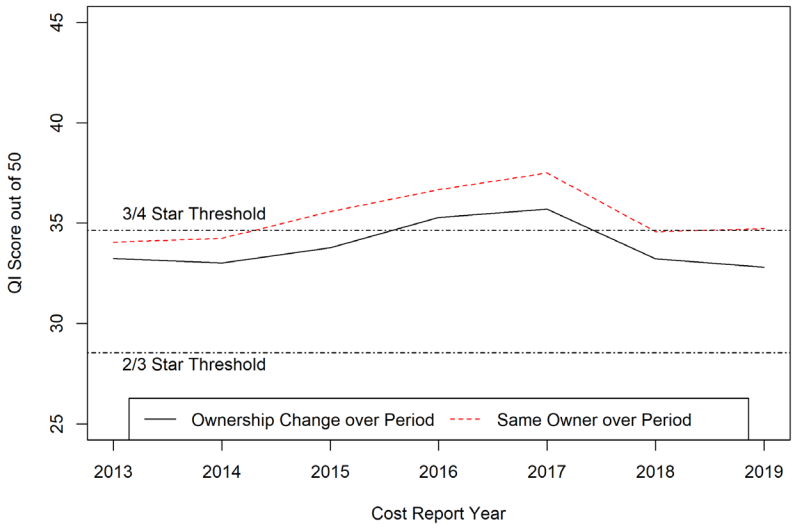


Figure 6: Quality Indicator Score Trajectories by Ownership Change Status (Constant Groups)



The Quality Indicator score tracks fairly closely between the two groups, with a 1.1 point increase in the same owner group relative to the CHOW group (diff-in-diff metric Figure 6). Most of this change appears to occur by 2015. The amount of annual gap due to including future CHOWs (Figure 7) is quite volatile: 2014- (-11%), 2015 - 82%, 2016 – (-36%), 2017 – 16%, 2018 – 44%. In 2016, future CHOWs (those sold in 2017-2019) had relatively higher QI scores than those who were already in the CHOW group (those sold between 2014-2016).

Figure 7: Quality Indicator Score Trajectories by Ownership Change Status (Cumulative CHOW Group)

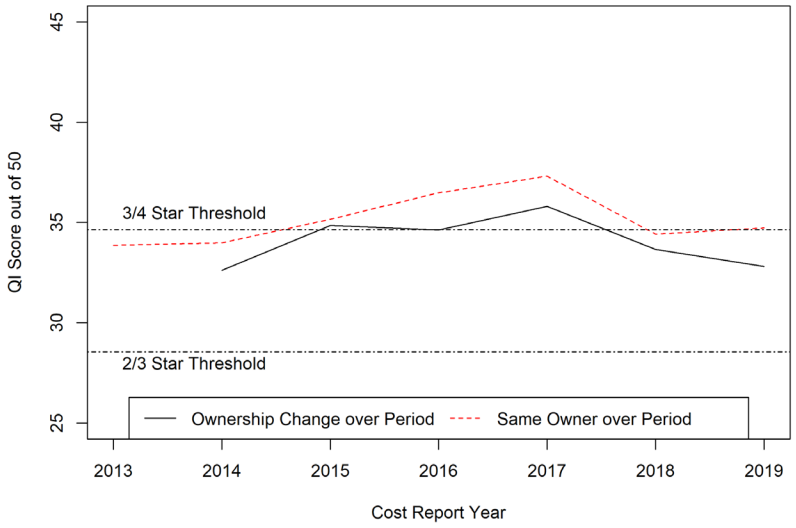
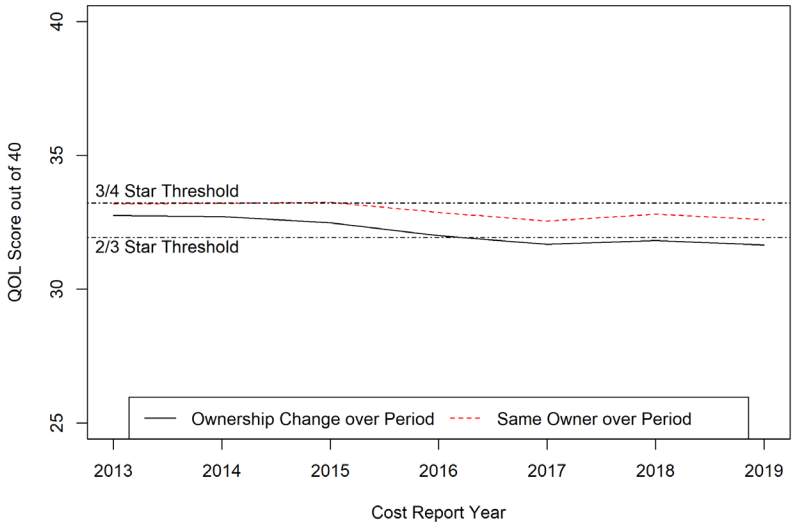


Figure 8: Quality of Life Score Trajectories by Ownership Change Status (Constant Groups)



Quality of life scores are down slightly over the period, but slightly less so for the same owner group (0.5 points higher relative to the CHOW group diff-in-diff metric Figure 8). The impact on the gap between the groups from including future CHOWs is concentrated in the early years: 2014 – 144%, 2015 - 69%, 2016 - 5%, 2017 – 5%, 2018 – 5%. There was essentially no average quality of life difference between the groups in 2014, and very little in 2015 (CHOWs occurring 2016-2019 had relatively worse QOL scores in 2014 and 2015 compared to CHOWs occurring in 2014-2015).

Figure 9: Quality of Life Score Trajectories by Ownership Change Status (Cumulative CHOW Group)

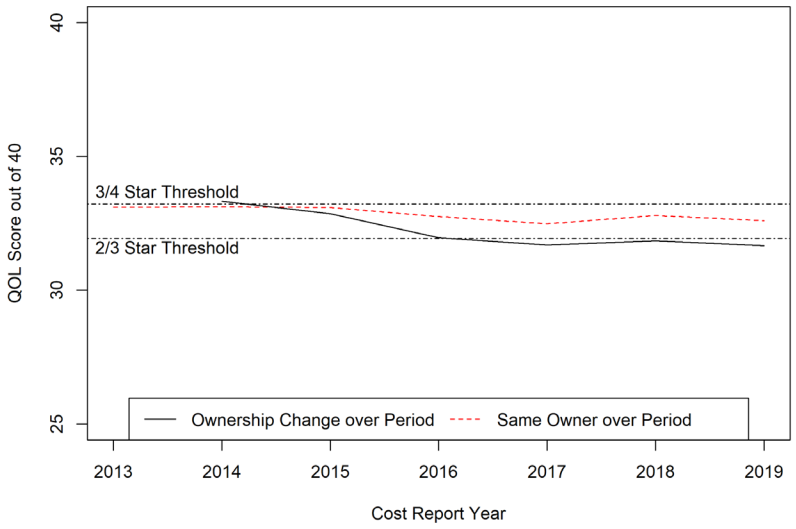
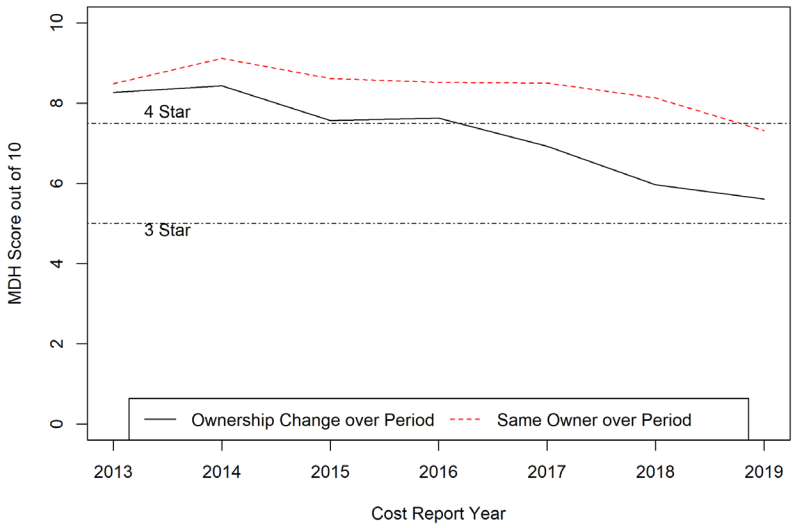


Figure 10: MDH Score Trajectories by Ownership Change Status (Constant Groups)



The same owner group had a 1.5 point increase in score on the 10 point Minnesota Department Health score over the CHOW group (Figure 10). The percentage of the gap due to including future CHOWs: 2014 – (-42%), 2015 - (-7%), 2016 – 14%, 2017 – 31%, 2018 – 16%. Chows occurring in 2016-2019 had relatively better MDH scores in 2014-15 than CHOWs occurring in 20154-15 (included later CHOWs reduced the gap in Figure 10).

Figure 11: MDH Score Trajectories by Ownership Change Status (Cumulative CHOW Group)

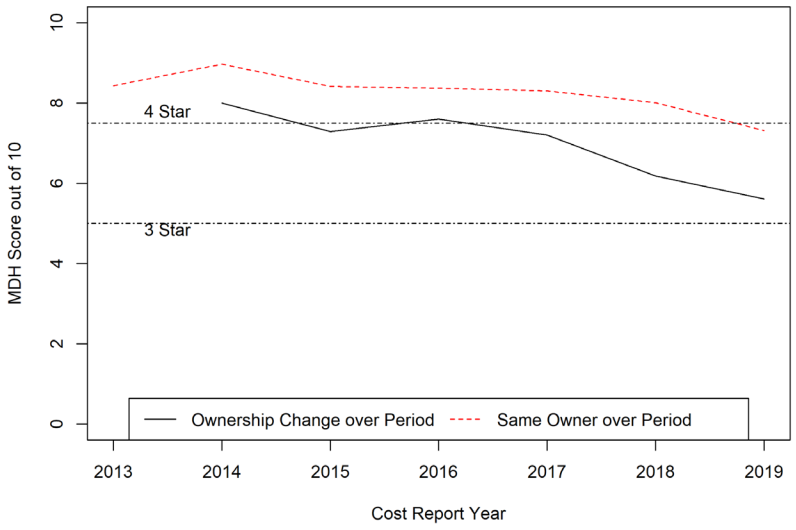
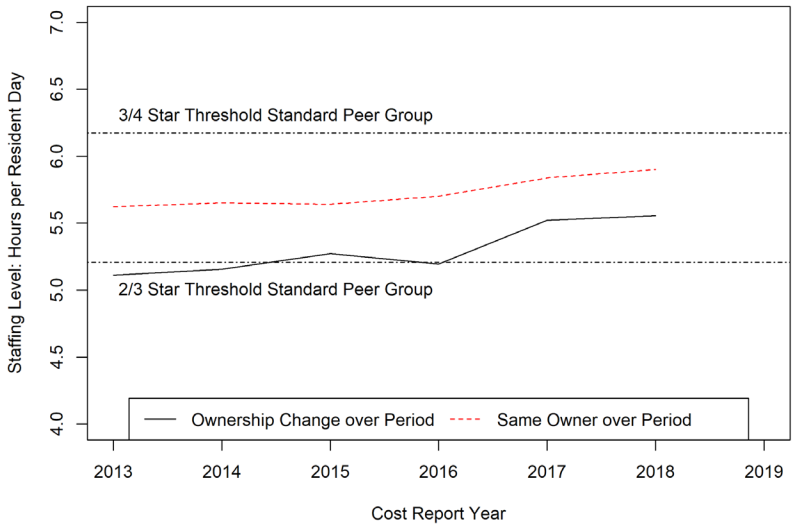


Figure 12: Staffing Score Trajectories by Ownership Change Status (Constant Groups)



The gap in staffing hours per resident day between the CHOW and same owner groups was reduced by 0.16 hours over the period (diff-in-diff metric). Much of the early gap in Figure 12 is attributable to the inclusion of future CHOWs, indicating that facilities going through a CHOW were more likely to have lower staffing levels. The percentage of the gap due to future CHOWs for each year: 2014 – 103%, 2015 – 81%, 2016 – 50%, 2017 – 23%, 2018 – 24%.

Figure 13: Staffing Score Trajectories by Ownership Change Status (Cumulative CHOW Group)

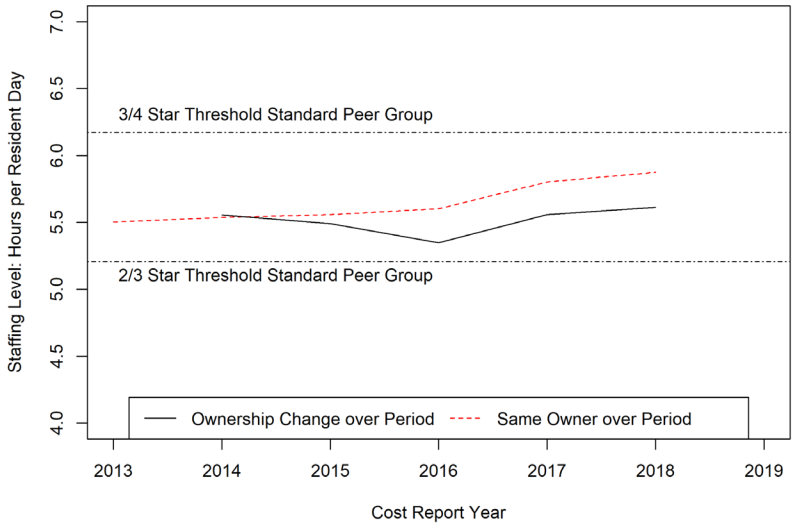
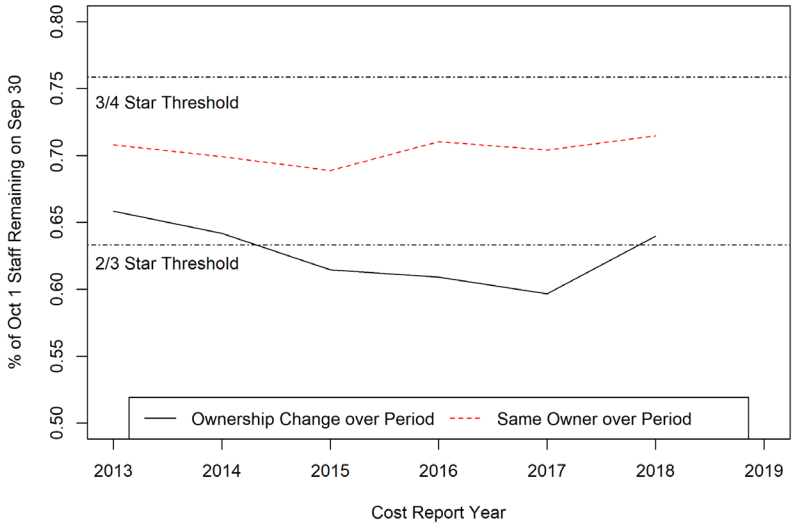


Figure 14: Retention Score Trajectories by Ownership Change Status (Constant Groups)



From 2013 to 2018, facilities experiencing a CHOW saw a 2.5% drop in retention relative to facilities with the same ownership (diff-in-diff metric). This figure would have been much larger in 2017, but the gap shrunk notably in 2018 (Figure 14). The percentage of the gap between groups due to the inclusion of future CHOWs: 2014 – 61%, 2015 – (-39%), 2016 – (38%), 2017 – (6%), 2018 – (-2%).

Figure 15: Retention Score Trajectories by Ownership Change Status (Cumulative CHOW Group)

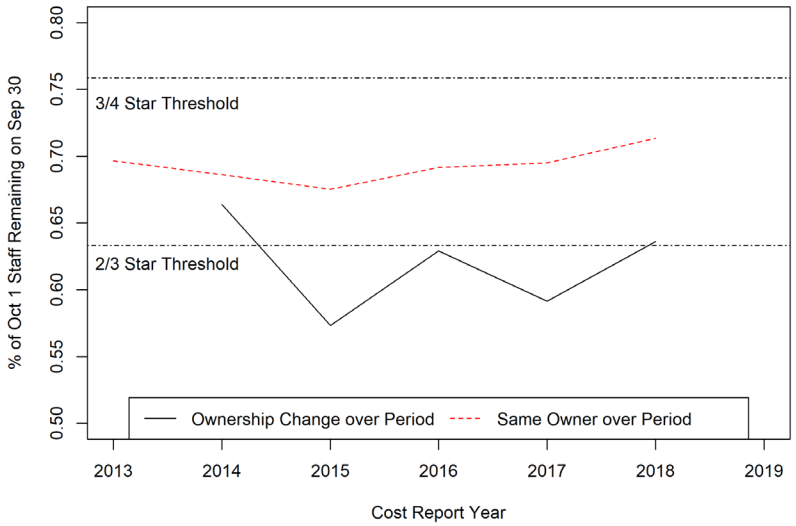
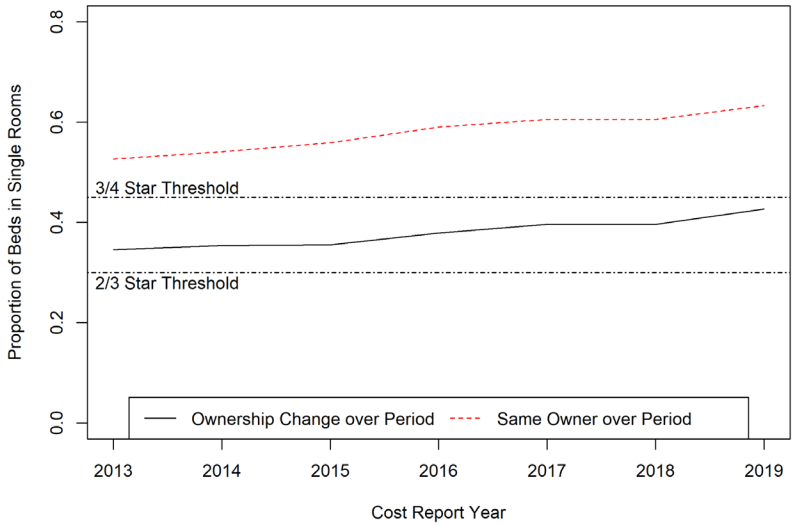


Figure 16: Private Room Score Trajectories by Ownership Change Status (Constant Groups)



The gap between same owner and CHOW facility proportion of private rooms widened by 2.5% over the period (diff-in-diff metric). Much of the visible gap in Figure 16 is due to facilities going through a CHOW in 2017. Percentage of gap due to future CHOWs: 2014 – 39%, 2015 – 54%, 2016 – 65%, 2017 – (-15%), 2018 – (-1%).

Figure 17: Private Room Score Trajectories by Ownership Change Status (Cumulative CHOW Group)

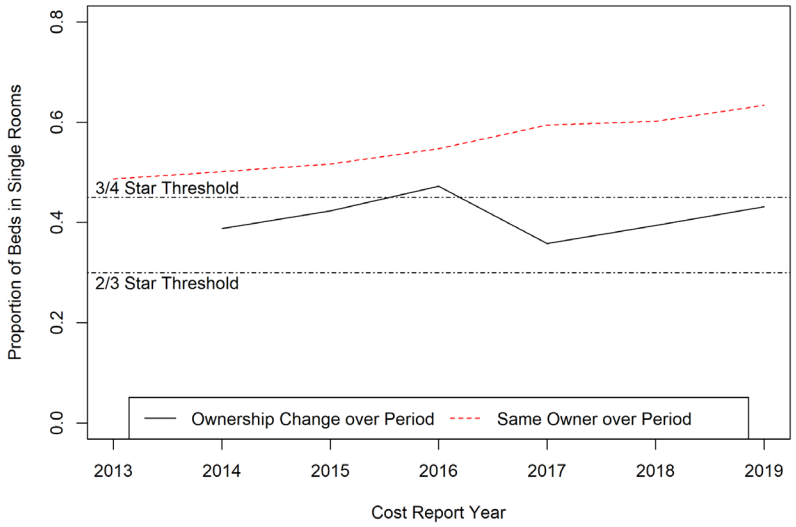
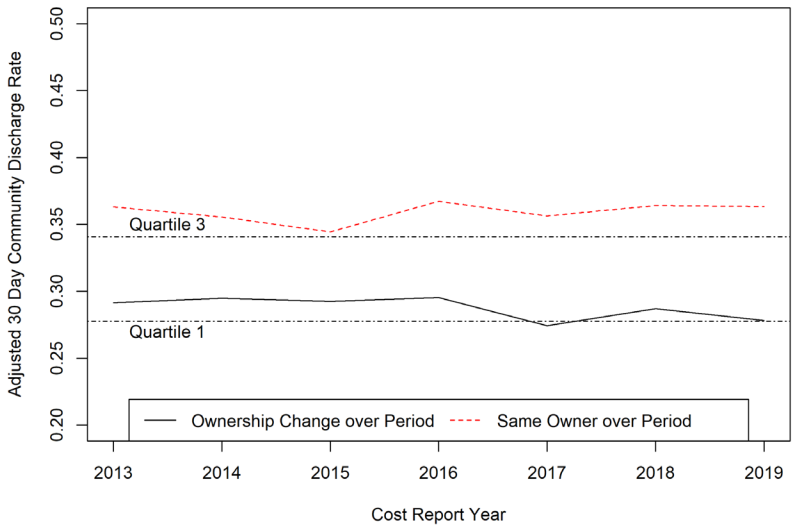


Figure 18: CD30 Score Trajectories by Ownership Change Status (Constant Groups)



The adjusted 30 day community discharge rate between the CHOW group and same owner group grew slightly over the period (1.2% increase, diff-in-diff metric). Initial CHOWs (2014) had much higher CD30 rates relative to later CHOWs. Percentage of gap due to future CHOWs: 2014 – 218%, 2015 – 1%, 2016 – 40%, 2017 – (-10%), 2018 – (-16%).

Figure 19: CD30 Score Trajectories by Ownership Change Status (Cumulative CHOW Groups)

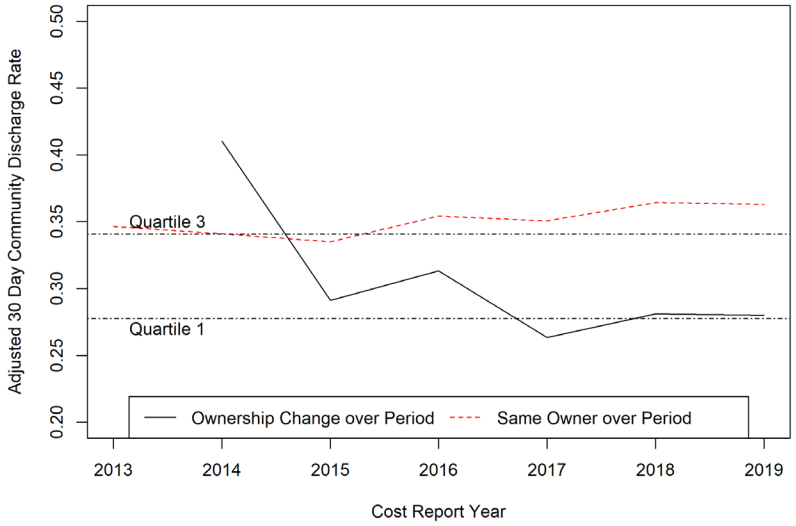
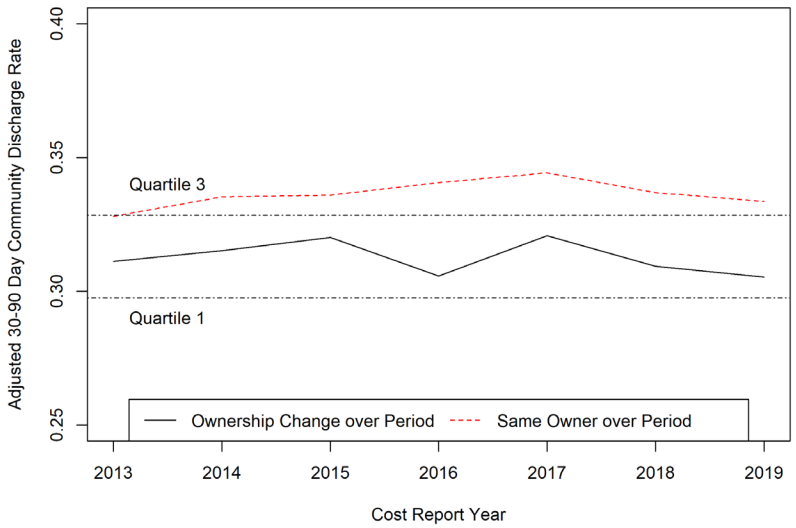


Figure 20: CD90 Score Trajectories by Ownership Change Status (Constant Groups)



The gap in adjusted 30-90 day community discharge rate between the CHOW group and same owner group grew slightly over the period (1.1% increase, diff-in-diff metric). Initial CHOWs (2014) had much higher CD90 rates relative to later CHOWs. Percentage of gap due to future CHOWs: 2014 – 113%, 2015 – (-42%), 2016 – 50%, 2017 – 8%, 2018 – (-15%).

Figure 21: CD90 Score Trajectories by Ownership Change Status (Cumulative CHOW Groups)

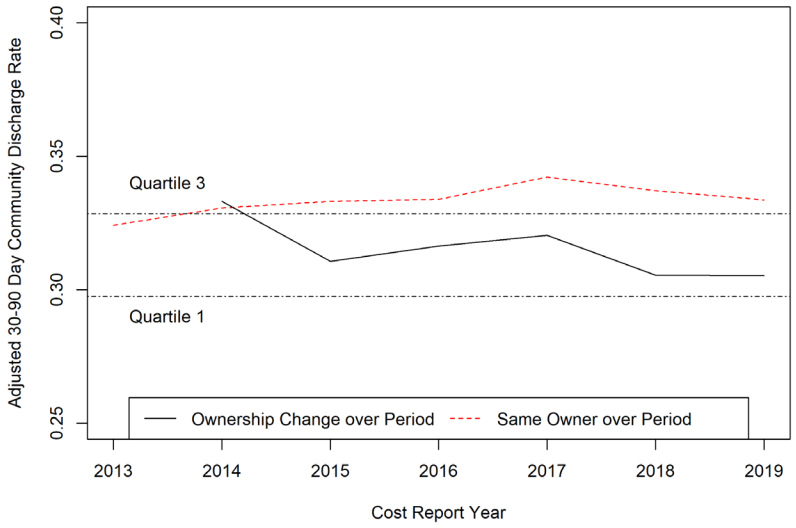
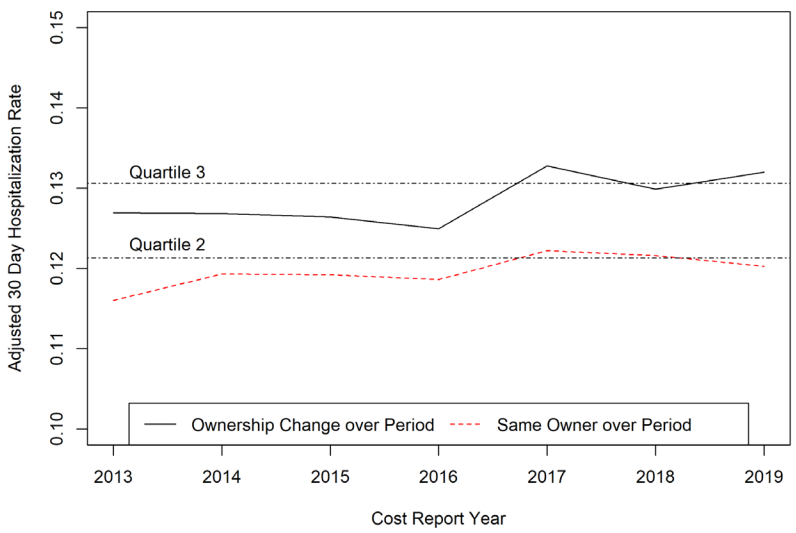


Figure 22: HRP Hospitalization Score Trajectories by Ownership Change Status (Constant Groups)



The gap in adjusted 30 day hospitalization rate between the CHOW group and same owner group did not change appreciably over the period (0.07% increase, diff-in-diff metric Figure 22). The impact of future CHOWs was highly volatile. Percentage of gap due to future CHOWs: 2014 – 72%, 2015 – (-47%), 2016 – 103%, 2017 – (-10%), 2018 – (-10%).

Figure 23: HRP Hospitalization Score Trajectories by Ownership Change Status (Cumulative CHOW Group)

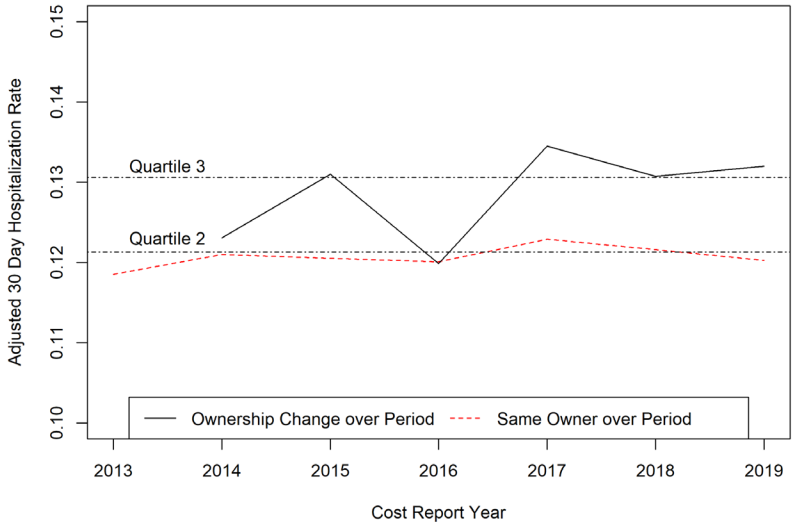
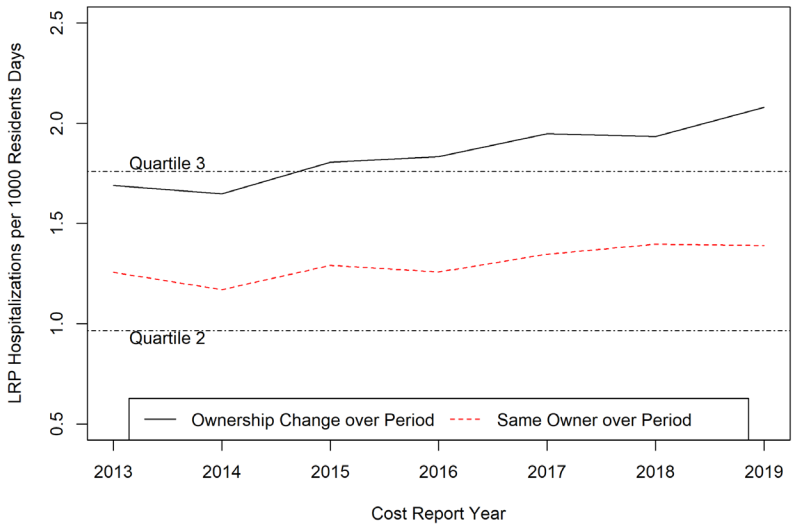
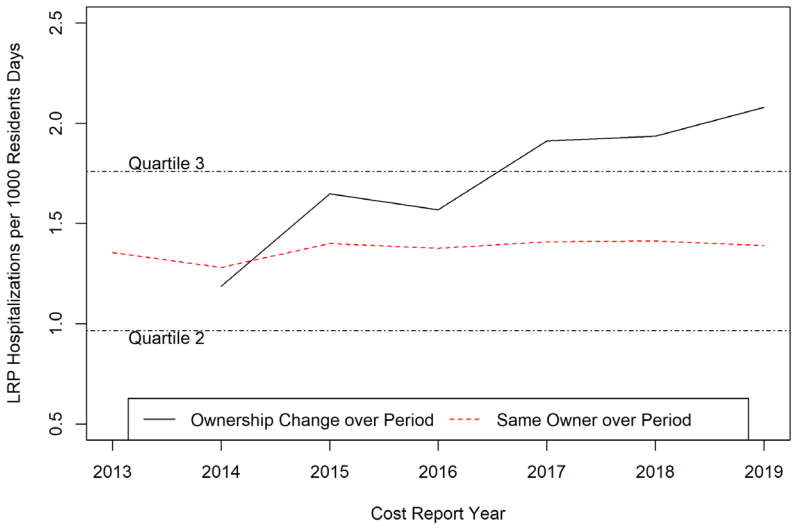


Figure 24: LRP Hospitalization Score Trajectories by Ownership Change Status (Constant Groups)



The gap in hospitalizations after 30 days per 1000 residents between the CHOW group and same owner group grew over the period by 1 hospitalization per 10,000 resident days (0.1 increase, diff-in-diff metric Figure 24). Later CHOWs (2017-2019) appear to contribute much of the low risk period hospitalization gap. Percentage of gap due to future CHOWs: 2014 – 120%, 2015 – 51%, 2016 – 69%, 2017 – 17%, 2018 – 2%.

Figure 25: LRP Hospitalization Score Trajectories by Ownership Change Status (Cumulative CHOW Group)



The following figures display overall quality scores for CHOW related facilities. Figure 26 displays the means by ownership groups (note that ownership names associated with groups 4-7 contained the same facilities and are represented by group 4, groups 9-12 contained 2020 CHOWs and so do not appear in these plots). Figure 27 - Figure 32 display individual facility quality scores within each ownership group. Since each line represents a single facility, trends are more volatile. However, comparing the solid line prior to the CHOW and to the dashed line following the CHOW can inform if new ownership if able to maintain or improve facility quality. A bolder dashed line has been added to each plot to display the loess smoothed ownership group trend and the individual circle give the overall CHOW group quality mean for each year. With the exception of ownership group 4 (CHOWs occur 2019 or later) and group 10 (single facility), all other ownership groups show a general decline in quality. For some groups this began prior to the CHOWs and continued after (groups 1 and 2) for others quality scores appear to have begun declining around the time of the CHOWs (groups 3 and 8). It is worth noting that not all facilities in the CHOW group appear to be experiencing declining quality scores, even within the same ownership group.

Figure 27 through Figure 32 display individual facilities, with each figure displaying a particular new ownership group's facilities. Each line displays a facility's quality scores over time, the solid portion of the line represents the pre-CHOW scores and the dashed portion of the line displays the cost year of the CHOW and subsequent years.

Groups size (number of facilities) is variable and impacts the volatility in the mean trend lines of Figure 26. Group size ranges from 1 (group 10) to 22 (group 3). The three largest groups (2 – 8 facilities, 3 – 22 facilities, and 8 – 13 facilities) begin the period with the highest relative quality means and end the period with an apparent downwards quality trajectory. The three smallest groups (1 – 4 facilities, 4 – 7 facilities, and 10 – 1 facility) appear to end with an upwards rebound in quality, although their means are more volatile over the period.

Figure 26: Mean Quality Scores by Ownership Group

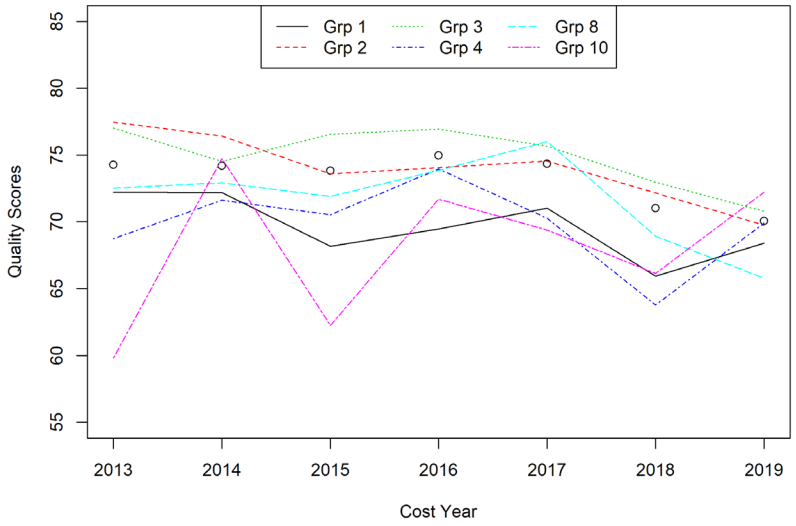


Figure 27 - Figure 32 display individual facility quality scores within each ownership group. Since each line represents a single facility, trends are more volatile. However, comparing the solid line prior to the CHOW and to the dashed line following the CHOW can inform if new ownership if able to maintain or improve facility quality. A bolder dashed line has been added to each plot to display the loess smoothed ownership group trend and the individual circle give the overall CHOW group quality mean for each year. With the exception of ownership group 4 (CHOWs occur 2019 or later) and group 10 (single facility), all other ownership groups show a general decline in quality. For some groups this began prior to the CHOWs and continued after (groups 1 and 2) for others quality scores appear to have begun declining around the time of the CHOWs (groups 3 and 8). It is worth noting that not all facilities in the CHOW group appear to be experiencing declining quality scores, even within the same ownership group.

Figure 27: Facility Quality Scores for Ownership Group 1

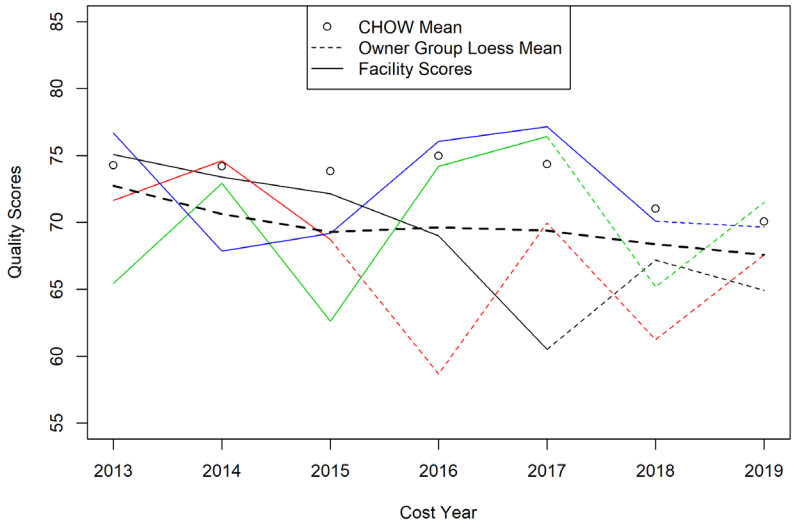


Figure 28: Facility Quality Scores for Ownership Group 2

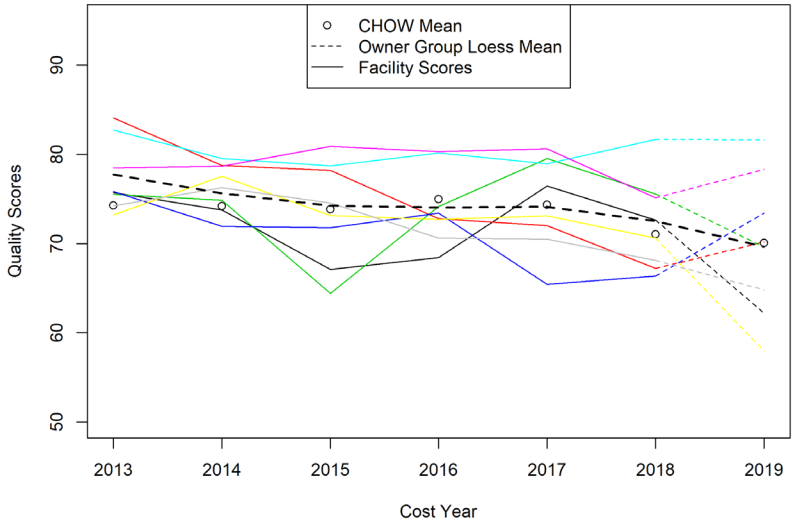


Figure 29: Facility Quality Scores for Ownership Group 3

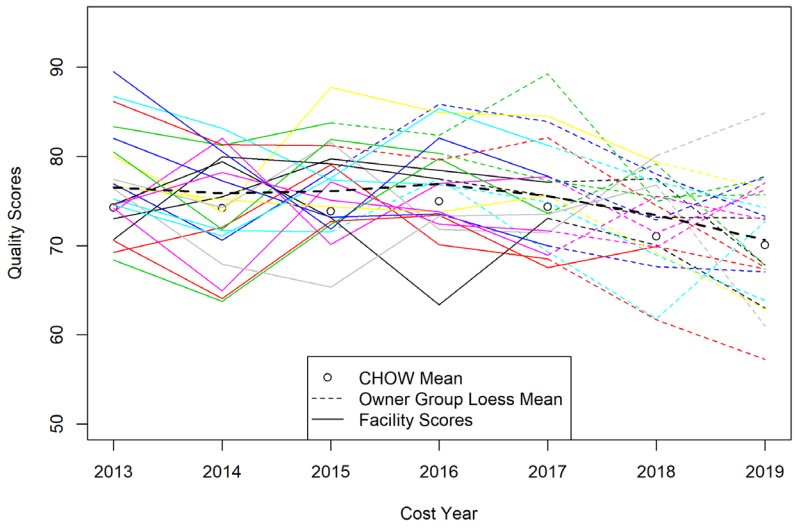


Figure 30: Facility Quality Scores for Ownership Group 4

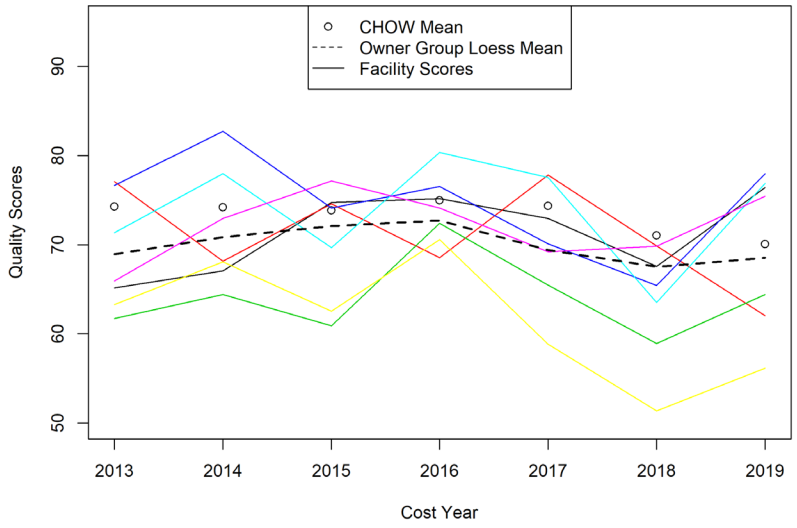


Figure 31: Facility Quality Scores for Ownership Group 8

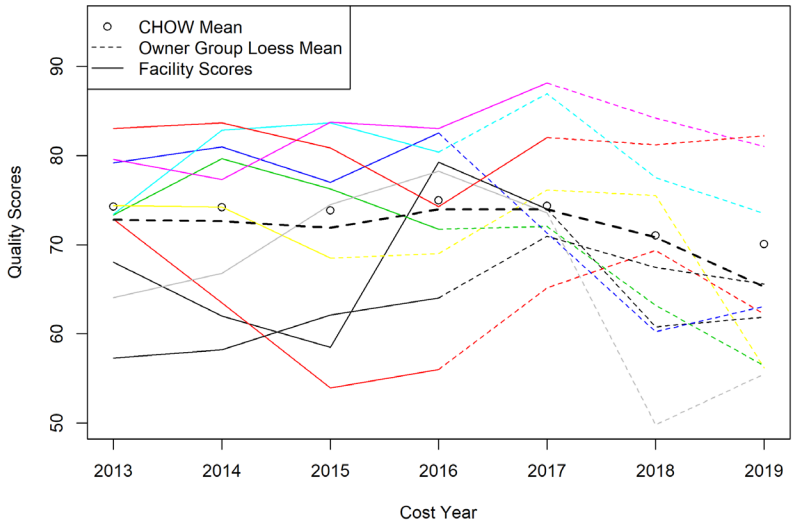
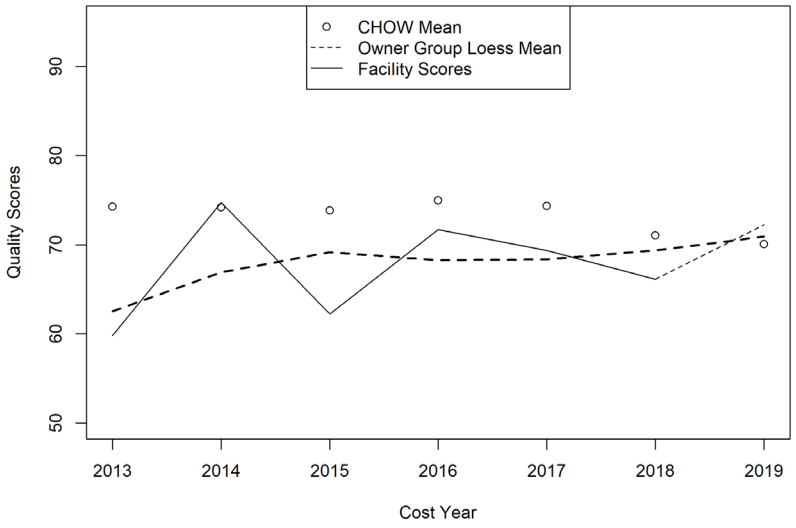


Figure 32: Facility Quality Scores for Ownership Group 10



The following tables (Table 7 and Table 8) present means for resident, staff, and quality variables in the year prior to and year following the CHOW cost year. Seventy four facilities were included in the CHOW group as having a cost year prior and post the CHOW cost year. This comparison is intended to highlight any changes that represent potential shift in operation related to the CHOW. A comparison group (non-CHOW) consists of all facilities that did not go through a CHOW in the following (for pre numbers) or previous (for post numbers) year. This will help differentiate changes due to CHOWs from general trends in the industry.

From Table 7, CHOWs experienced an 11% drop in admissions in the cost year after the CHOW year relative to the cost year prior to the CHOW year. In comparison, facilities with constant ownership over the three year time frame experienced an increase in annual admission of 4%. The year following a CHOW, the CHOW group saw a slight increase in staff retention driven primarily by 16% increase in Nurse Administrator retention rates and 7% increase in RN retention rates. Conversely, social worker retention rates in CHOWs dropped from 63 to 57% (-10% change). Despite the slight improvement, CHOW retention rates still lag substantively behind constant ownership facilities (Figure 14). CHOW facilities increased salary per resident day of Nurse Administrators (27%) and Social Workers (33%) by more than the non-CHOW group, but RN salary increases lagged. CHOWs appear to be making cuts to laundry expenses, dental and pension benefits, while increasing medical and scholarship benefits at a slower rate than peers, and increasing administrative management fees at a higher rate (40% increase vs. 13%).

Table 7: Staffing and Resident Profile Before and After the Change of Ownership

	Pre CHOW	Post CHOW	CHOW Change	Non-CHOW Change
Annual Admissions	**201**	**178**	**-11%**	**4%**
Medicaid Revenue	**$ 2,876,475**	**$ 3,114,980**	**8%**	**11%**
Medicaid Resident Days	**14,199**	**13,424**	**-5%**	**-4%**
Total Resident Days	**23,103**	**21,623**	**-6%**	**-4%**
Acuity	**1.02**	**1.02**	**0%**	**0%**
Total Retention	**60%**	**61%**	**1%**	**0%**
Retention: Activities Staff	**68%**	**69%**	**3%**	**-2%**
Retention: CNA	**57%**	**58%**	**1%**	**0%**
Retention: LPN	**66%**	**65%**	**-2%**	**0%**
Retention: NA	**61%**	**71%**	**16%**	**-2%**
Retention: ODC	**17%**	**10%**	**-41%**	**-2%**
Retention: RN	**57%**	**61%**	**7%**	**1%**
Retention: Social Worker	**63%**	**57%**	**-10%**	**-1%**
Retention: TMA	**31%**	**32%**	**4%**	**3%**
Retention: CNA/TMA	**57%**	**58%**	**2%**	**0%**
Nurse Administrator Salary PRD	**$9.27**	**$11.80**	**27%**	**16%**
RN Salary PRD	**$18.92**	**$20.43**	**8%**	**19%**

	Pre CHOW	Post CHOW	CHOW Change	Non-CHOW Change
LPN Salary PRD	$19.22	$18.55	-3%	7%
CNA Salary PRD	$33.66	$37.35	11%	13%
TMA Salary PRD	$2.89	$3.22	11%	24%
DC Trainer Salary PRD	$0.70	$0.41	-42%	20%
Medical Records Salary PRD	$2.85	$3.25	14%	12%
Social Worker Salary PRD	$2.90	$3.85	33%	14%
Activities Staff Salary PRD	3.64	4.08	12%	12%
Other DC Staff Salary PRD	0.22	0.35	60%	-1%
Therapy Salary PRD	$0.06	$-	-100%	-58%
Other Care Staff Salary PRD	$9.68	$11.52	19%	11%
Dietary Total Cost PRD	$13.75	$15.11	10%	8%
Laundry Total Cost PRD	$4.14	$3.77	-9%	7%
Housekeeping Total Cost PRD	$6.83	$7.59	11%	9%
Plant Total Cost PRD	$12.26	$13.80	13%	9%
Admin Total Cost PRD	$34.73	$36.63	5%	11%
Dental Cost PRD	$0.32	$0.09	-72%	13%
Pension Cost PRD	$0.55	$0.54	-2%	16%
Admin Management Fees PRD	$8.00	$11.20	40%	13%
Scholarship Cost PRD	$0.32	$0.50	58%	71%
Group Medical PRD	$7.90	$8.30	5%	25%
LPN DC Hours PRD	0.79	0.73	-7%	-5%
RN DC Hours PRD	0.60	0.62	4%	9%
CNA DC Hours PRD	2.24	2.27	1%	1%
Licensed DC Hours PRD	1.39	1.35	-2%	1%
Total DC Hours PRD	4.48	4.62	3%	3%
Activities DC Hours PRD	0.23	0.25	7%	5%

	Pre CHOW	Post CHOW	CHOW Change	Non-CHOW Change
Nurse Administrator DC Hours PRD	0.30	0.35	15%	9%
Social Worker DC Hours PRD	0.13	0.16	23%	6%
TMA DC Hours PRD	0.16	0.19	17%	12%
Other Direct Care Hours PRD	0.02	0.02	30%	-8%

***Pre numbers come from the cost report prior to the sale cost year and post numbers come from the cost report following the sale cost year (i.e. For a facility occurring in 2014 cost year, the 2013 cost report is used for the pre-numbers and the 2015 cost report is used for the post-numbers). CHOW change is the percentage change in the post period from the pre-period. Non-CHOW change is the percent change in the post period from the pre-period for all facilities that did not go through a CHOW in the following (for pre numbers) or previous (for post numbers) year.**

Table 8 displays changes in quality scores following the same format as Table 7. The most notable change in CHOW facilities relative to the comparison group were a dip in adjusted 30 day community discharge rates, a two point drop in quality indicator scores, and a drop of a point in MDH inspection scores.

Table 8: Quality Scores Before and After CHOW Event

	Pre CHOW	Post CHOW	CHOW Change	Non-CHOW Change
Adjusted 30 Rate CD Rate	0.31	0.28	-9%	1%
Adjusted 31-90 Day CD Rate	0.31	0.31	-1%	1%
Adjusted 30 Day Hospitalization Rate	0.13	0.13	3%	1%
Unadjusted Low Risk Period Hospitalizations per 1000 Resident Days	1.90	2.04	7%	8%
VBR Quality Score (Old Version)	75.03	72.42	-3%	-1%
Quality Indicator Score (50 Points)	35.36	33.54	-5%	0%
MDH Score (10 Points)	7.64	6.79	-11%	-6%
Quality of Life Score (40 Points)	32.26	31.88	-1%	-1%
Staffing Score	5.42	5.61	3%	2%
Case Mix	1.02	1.01	-1%	0%
Staff Retention Score	0.60	0.60	0%	0%
Private Room Score	0.38	0.40	4%	6%

***Pre numbers come from the cost report prior to the sale cost year and post numbers come from the cost report following the sale cost year (i.e. For a facility occurring in 2014 cost year, the 2013 cost report is used for the pre-numbers and the 2015 cost report is used for the post-numbers). CHOW change is the percentage change in the post period from the pre-period. Non-CHOW change is the percent change in the post period from the pre-period for all facilities that did not go through a CHOW in the following (for pre numbers) or previous (for post numbers) year.**