B | Governance Studies
at BROOKINGS

DECEMBER 2016

Five-star ratings for sub-par service: Evidence of inflation in nursing home ratings

Xu Han, Niam Yaraghi, Ram Gopal

Xu Han is a Ph.D. Candidate at the Department of Operations and Information Management at the University of Connecticut's School of Business.

Niam Yaraghi is a Fellow in the Brookings Institution's Center for Technology Innovation. He is an expert on the economics of health information technologies.

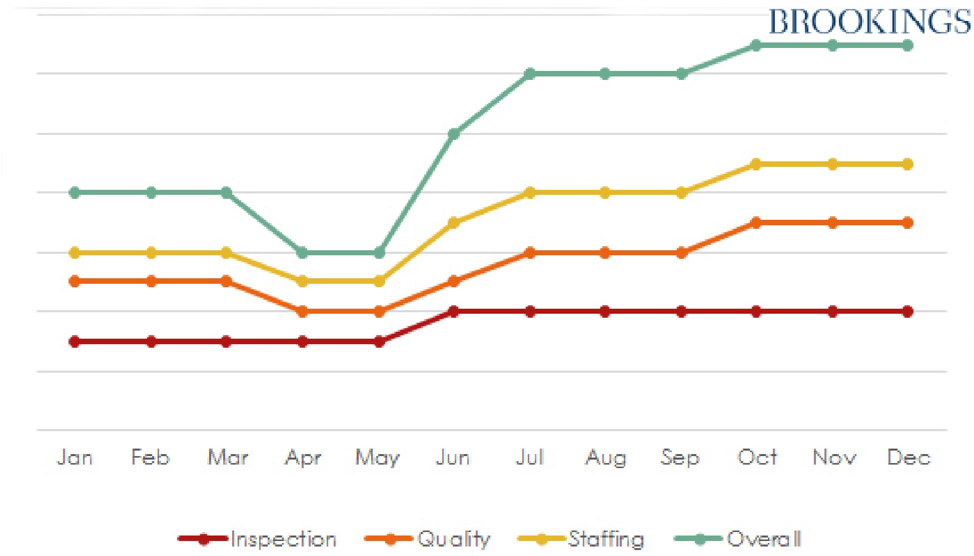
Ram Gopal is the GE Capital Endowed Professor of Business and the Chair of Operations and Information Management Department at University of Connecticut's School of Business.

INTRODUCTION

Nearly two million Americans spend an average of 835 days of their lives in one of the 15,700 nursing home facilities in the United States (National Center for Health Statistics 2009). The Department of Health and Human Services estimates that in 2009, 4.1 percent of Americans over 65 years old lived in these facilities. This percentage increases with age, ranging from 1.1 percent in the population of 65 to 74-year-olds to 13.2 percent in the population older than 85 (Fowles 2012). In 2012 alone, Medicaid spent $140 billion on long-term services and supports (Eiken et al. 2014). Despite the importance of nursing homes in the quality of life of millions of Americans and the billions of dollars spent on them, very little information has been available about their service quality. The Centers for Medicare & Medicaid Services (CMS) designed and implemented its nursing home rating system after a congressional hearing in 2007 in which Senator Ron Wyden asked why it was "easier to shop for washing machines than it is to select a nursing home" (Duhigg 2007). Given the lack of alternative information resources on nursing homes, the publically available CMS rating has become the gold standard in the industry since its inception, and has been widely popular among patients, physicians, and payers (Thomas 2014). The recent study of Werner et al. (2016) sheds light on the importance of CMS ratings for nursing homes; according to their analysis, after the release of the ratings, the market share of one-star facilities decreased by eight percent while the market share of five-star facilities increased by more than six percent. Given its important role, nursing homes have a significant incentive to improve their ratings; however, these ratings may not always reflect true quality. As we discuss below, the rating system is prone to inflation by nursing homes.

The CMS rating system is based on three domains: on-site inspection, staffing, and quality measures. While independent, CMS-certified inspectors conduct and report on the on-site inspections, the other two domains are self-reported by nursing homes. CMS first assigns an initial star rating to all nursing homes based on their annual on-site inspection results.[1] Nursing homes are then assigned star ratings for the staffing[2] and quality measures[3] domains. The overall star rating is then calculated by considering the on-site inspection rating as the baseline, adding one star if any self-reported domain is five stars and subtracting one star if any self-reported domain is one star.[4] An example is provided in Table 1 and Figure 1 to demonstrate the rating dynamics and the corresponding events for a randomly selected nursing home in 2009.

Figure 1. The graphical representation of a nursing home's rating dynamics



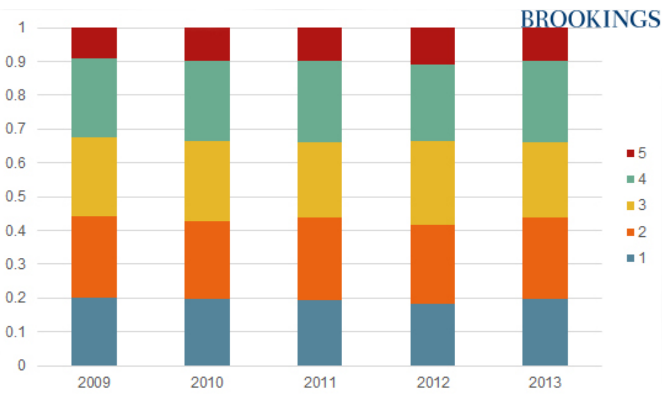
Notes: [a] In the first quarter of 2009, the nursing home received 3 stars in inspection. It reports 2 stars in quality measures and 1 star in staffing. The resulting overall rating is 2 stars.
[b] In April, the reported quality measure reduces to 1 star, with the other two domains unchanged. As a result, the overall rating reduces to 1 star.
[c] In June, a new inspection is conducted, in which the nursing home receives 4 stars. The staffing data is also reported together with the inspection in June to be 2 stars. The resulting overall rating is 3 stars.
[d] In July, the quality measures are newly reported to be 2 stars. With the other domains unchanged, the overall rating increases to 4 stars, since none of the self-reported domains are 1 star.
[e] In October, the quality measures are newly reported to be 3 stars. This change, however, does not affect the overall rating.

Table 1. An example of a nursing home's rating dynamics

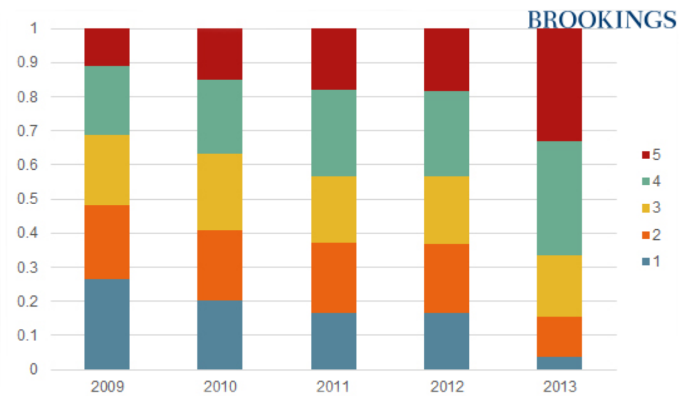
Month	*Overall*	*Inspection*	*Quality Measurement*	*Staffing*
January	2	3	2	1
February	2	3	2	1
March	2	3	2	1
April	1	3	1	1
May	1	3	1	1
June	3	4	1	2
July	4	4	2	2
August	4	4	2	2
September	4	4	2	2
October	4	4	3	2
November	4	4	3	2
December	4	4	3	2

The two self-reported domains can fundamentally change a nursing home's overall rating. For example, it is possible for an average nursing home that has received three stars in the on-site inspection to gain two additional stars based on self-reported measures and become an excellent five-star nursing home. As a result, the overall rating can be quite different from the on-site inspection rating. Figure 2 shows how the ratings in each of these domains have shifted to higher stars during a period of five years from 2009 to 2013. The proportion[5] of on-site inspection star ratings remains unchanged in the five years, as shown in Figure 2(a). However, the number of nursing homes that claim high performance in the self-reported domains has continuously increased over the past five years. As shown in Figure 2(b), in 2009, about 40 percent of nursing homes self-reported to be four or five stars in the quality measures domain. This percentage increased to 60 percent in 2013. On the other hand, about 20 percent of nursing homes self-reported to be one-star in 2009, but less than 10 percent of nursing homes self-reported to be one-star in 2013. In the staffing domain, the number of highly rated nursing homes also significantly increased over this period, as shown in Figure 2(c). Consequently, the overall rating is consistently skewed to the higher end over time. As shown in Figure 2(d), the portion of four- or five-star nursing homes increased from 35 percent to 55 percent over the five years.

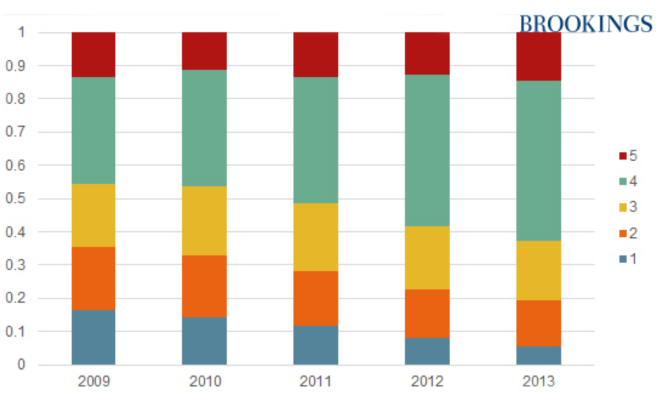
Figure 2. Distribution of nursing home ratings from 2009 to 2013



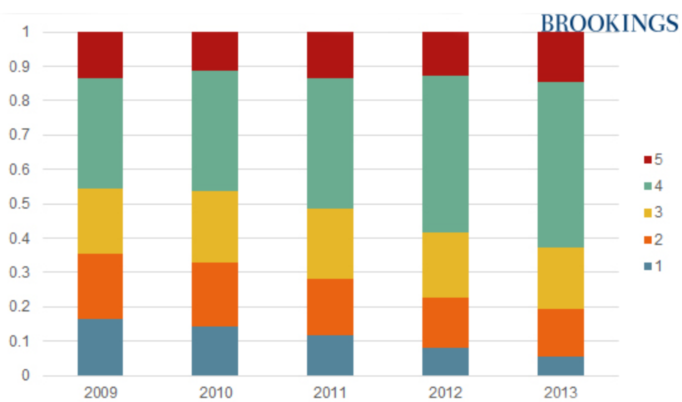
(a) On-site inspection



(b) Quality Measure



(c) Staffing



(d) Overall rating

Note: Colors represent different star rating groups

The trend we observe in Figure 2 can be interpreted in two ways: On one hand, supporters can argue that increased levels of self-reported measures are genuine and represent an honest effort by nursing homes to constantly improve their services. On the other hand, however, skeptics may argue that the improved ratings are not legitimate but are rather a result of nursing homes' success in developing strategies to manipulate the system and inflate their ratings. Cases have been reported in which patients' personal experiences differ significantly from the star ratings. Some highly-rated nursing homes are sued for substandard care, even causing death of patients due to improper medical treatments (Thomas 2014).

Since late 2014, CMS has gradually announced new policies to improve the nursing homes' rating system (Medicare 2016). These corrective policies include the expansion of targeted surveys, including additional measures in the quality measures domain, and adding payroll information to the staffing reports. Despite these notable improvements, the structure of the rating system has not been changed and it still heavily relies on the self-reported domains—thus, the newly revised system continues to be prone to manipulation by false self-reported measures. Since it is not clear whether the continued increase in ratings of self-reported domains is driven by nursing homes' legitimate efforts to constantly improve their services or it is rather a signal of rating inflation and fraudulent self-reporting, the objective of this research is to answer this question and investigate the existence and the extent of inflation in the nursing homes' rating system.

It is not clear whether the continued increase in ratings of self-reported domains is driven by nursing homes' legitimate efforts to constantly improve their services or it is rather a signal of rating inflation and fraudulent self-reporting. The objective of this research is to answer this question and investigate the existence and the extent of inflation in the nursing homes' rating system.

This research is based on the publicly available data provided by multiple government agencies including CMS, California's Office of Statewide Health Planning and Development (OSHPD), and the California Department of Public Health (CDPH). Our empirical strategy consists of four steps as discussed below.

First, we explore the financial incentives for nursing homes to improve their star ratings using a combination of CMS rating data and OSHPD financial data. We find a significant positive association between the change in star ratings and the financial incentives. That is, nursing homes with higher financial incentives are more likely to improve star ratings after self-reporting.

Second, to prove the existence of rating inflation, we initially analyze the correlation between the CMS inspection and nursing homes' self-reported results. If the self-reported improvement is legitimate, we expect it to be reflected in the inspection results of the subsequent period. We also expect the CMS inspection rating and self-reported ratings within the same year to be closely associated. Our correlation analysis results, however, show almost no correlation between the inspection and self-reported results, and shed doubt on the legitimacy of self-reported measures. We then further corroborate the results of our correlation analysis by examining additional data on patient complaints provided by CDPH: If we assume that the ratings are not inflated, then we should observe similar service qualities among the nursing homes with similar *overall* ratings. Moreover, we should observe increased service quality among the nursing homes that initially had the same *inspection* rating but ended up with a higher *overall* rating as a result of their high self-reported measures. Our results, however, show significant differences between the service qualities of the nursing homes with the same *overall* rating. Moreover, no significant difference exists in the service quality of nursing homes with the same *inspection* rating. This result serves as strong evidence of the existence of inflation in the current rating system, as it points to the fact that the service quality is predicted by the on-site inspection ratings, which cannot be inflated, rather than the overall ratings, which can be inflated.

Third, to estimate the extent of rating inflation, we develop a prediction model and apply it to estimates of the proportion of nursing homes that have inflated their self-reported ratings. By using a 95 percent confidence interval, we identify around six percent of nursing homes in the suspect population to be likely inflators in the current system.

Fourth, we conduct a variable importance analysis to classify the factors that their change contributes the most to the probability of being an inflator. Our results demonstrate the shortcomings of the current rating system and call for significant reforms in how CMS and other payers evaluate the quality of nursing homes.

Our results demonstrate the shortcomings of the current rating system and call for significant reforms in how CMS and other payers evaluate the quality of nursing homes.

The paper proceeds as follows. In section two, we describe our data collection procedure and explore the underlying financial incentives for nursing homes to improve their ratings. In section three, we first perform correlation analysis between the CMS-conducted inspection and self-reported measures, which cast doubt on the existence of rating inflation. We then demonstrate our conclusion by performing a more rigorous complaint-based analysis. A prediction model is developed in section four to identify likely rating inflators and evaluate the performance of the system. A variable importance analysis is then conducted to show key characteristics of the inflators. We conclude the whole paper in section five, by discussing the limitations of the existing work and describing necessary future work.

DATA COLLECTION AND FINANCIAL INCENTIVE ANALYSIS

DATA COLLECTION

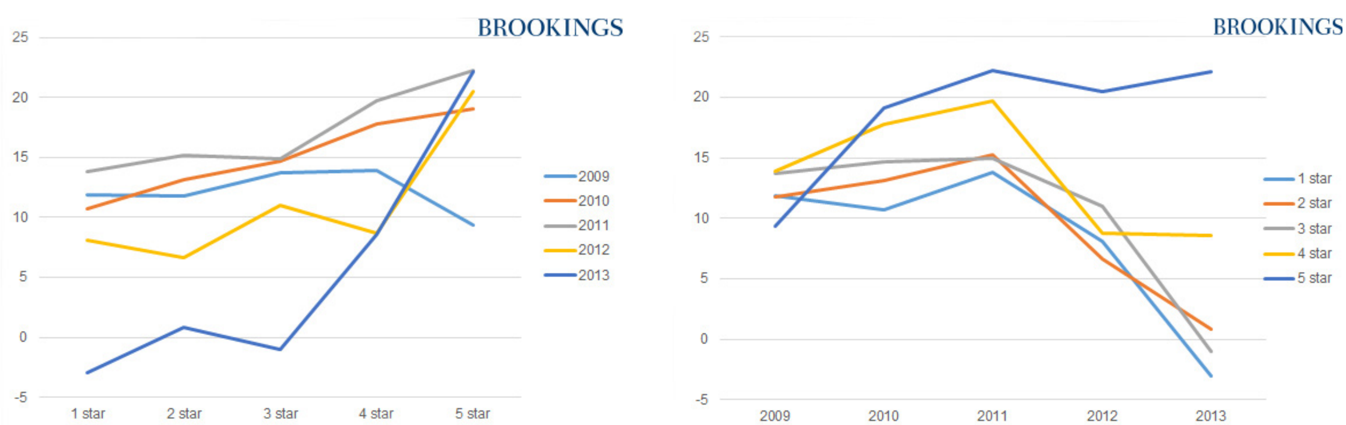
Our analysis is based on publicly available datasets from three sources: CMS, OSHPD, and CDPH. The CMS dataset includes performance details on each of the criteria used within the three domains of inspection, staffing, and quality measures. For each nursing home, these detailed metrics are accompanied by the corresponding star rating in the three domains as well as the overall star rating. This dataset also includes other descriptive details for nursing homes such as location, size, certification, ownership information, and council type. The pooled dataset consists of records from 1,219 nursing homes in the state of California over a five year period from 2009 to 2013. The OSHPD data includes detailed financial information on California nursing homes over the same period of time. In this dataset, nursing homes' source of revenue is categorized into health care and non-health care sections. The health care section is further classified by revenue source into Medicare, Medicaid, and self-paying. The corresponding revenue and expense details for each section are provided, and the profits can be easily calculated. The CDPH data contain detailed patient complaints, facility self-reported incidents, and inspection results for California nursing homes. The CDPH complaints data not only covers complaints that CMS has already considered in its rating procedure, but also includes patients' complaints and deficiency reports that are only available at the state level and are not reported to CMS.

NURSING HOMES’ FINANCIAL INCENTIVES

Observed rating improvements consist of both legitimate efforts and self-reporting inflation. In order to demonstrate the existence of rating inflation, we should show that the rating increase is beyond a range which can be explained by legitimate efforts. In our model, we perform a financial incentive analysis to establish the connection between a nursing home’s financial incentive and the increase in its star rating. We then show that this increase is far beyond the limit which can be explained by legitimate efforts.

We combine the CMS rating data and OSHPD financial data to demonstrate the financial implications of star ratings for nursing homes. The average profit per-day per-patient is calculated for nursing homes in each overall rating group, as shown in Table 2. These averages serve as an estimate of the daily profit that a nursing home can expect per-resident for the corresponding overall rating. The difference is significant. For example, a nursing home that receives three stars in on-site inspection may only expect a $10.44 profit from treating one patient for one day. However, if it gains two additional stars after self-reporting and achieves an overall rating of five stars, its expected profit increases to $16.88. Figure 3 shows the profit trend for each of the star rating groups over the five years.[6] The results demonstrate nursing homes’ incentives to achieve the highest possible ratings from the financial perspective, and provide a quantifiable metric to measure such incentives.

Figure 3. Profit trend over the period of 2009-2013



(a) Using Star Ratings as the Horizontal Axis *(b) Using Years as the Horizontal Axis*

In our model, we define the financial incentive of a nursing home to be the profit difference between its inspection rating and the highest overall rating it could potentially obtain after self-reporting, as shown in Table 2. Note that the financial incentive arises from the expectations in both profits and losses. It is possible for a nursing home that has received five stars from the on-site inspection to lose two stars if it receives one star rating in the self-reported domains. However, it is very unlikely that a nursing home with a perfect on-site inspection is significantly under-staffed or provides very poor quality of care. In our dataset, while 125 nursing homes initially rated three stars in inspection gained two additional stars after self-reporting, only four nursing homes initially rated five stars in inspection lost two stars after self-reporting.

Table 2. Definition of financial incentive

On-site inspection rating	Expected profit [a]	Maximum possible overall rating	Maximum expected profit [b]	Financial Incentive [c]
5	16.879	5	16.879	0
4	13.246	5	16.879	3.633 (Level 5– Level 4)
3	10.438	5	16.879	6.441 (Level 5– Level 3)
2	9.234	4	13.246	4.012 (Level 4– Level 2)
1	8.518	2	9.234	0.716 (Level 2– Level 1)

Notes: [a] If inspection rating unchanged. The expected profit is the average per patient per day profit for the corresponding star rating group.

[b] If maximum possible overall rating realized.

[c] Difference between expected profit and expected loss.

EMPIRICAL MODEL SPECIFICATION

We focus on changes in star ratings that happen because of the self-reported measures. Our dependent variable, *StarChange,* is equal to the difference between the overall rating and the on-site inspection rating. For example, if the nursing home receives three stars from on-site inspection but receives a five-star overall rating after including its self-reported measures on staffing and quality measure domains, then the *StarChange* would be equal to two.

By definition, *StarChange* can only take discrete values of 2, 1, 0, -1 and -2, and thus we use an ordinal logistic specification in which *StarChange* is modeled as a function of a vector of independent variables. *StarChange* is determined by a set of parameters, $\alpha_{-2}, \alpha_{-1}, \alpha_0, \alpha_1$, which define the cutoff points of the five levels. *StarChange* for nursing home at year *t* can be modeled as follows where $j \in \{-2,-1,0,1\}$ and $\boldsymbol{x}$ is a vector of the following independent variables: *Incentive*, *Competition*, *ResTotal*, *ForProfit*, *Medicare*, *Medicaid*, *ResCouncil* and *FamCouncil*.

$$P(StarChange_{it} \leq j) = \frac{\exp\left(\alpha_j + x_{it}' \beta\right)}{1+\exp\left(\alpha_j + x_{it}' \beta\right)} \qquad (1)$$

Among the independent variables, the main effect we consider in our model is the nursing homes' financial incentive, denoted by *Incentive,* and as shown in Table 2, varies over time depending on the inspection rating of a nursing home. The variable *Competition* describes the number of competing nursing homes in a 10-mile radius. The capacity of residential patients in each nursing home represents its size, and is denoted by *ResTotal*. Variable *ForProfit* defines a nursing home's ownership type and is equal to one if the nursing home is for-profit and zero otherwise. Variables *Medicare* and *Medicaid* define a nursing home's certification. *Medicare* is equal to one if the nursing home is Medicare certified, likewise, *Medicaid* is equal to one if the nursing home is Medicaid certified. By law, nursing homes are required to allow councils set up by residents or their family members. These councils facilitate the communication with staff and get problems resolved more efficiently. Since nursing home residents may be more vulnerable than other people due to their health conditions, the residential council and family council can function very differently in resolving issues and handling complaints. In our model, binary variables *ResCouncil* and *FamCouncil* are included to respectively, denote the council types as residential and family. A nursing home can have both types of councils. Table 3 provides the summary statistics of all variables in our model.

Table 3. Variable summary statistics

Variables	Mean	Standard Deviation	Minimum	Maximum
Incentive	3.443	2.178	0	6.441
Competition	41.066	37.883	0	150
ResTotal	86.985	43.648	12	373
ForProfit	0.894	0.308	0	1
Medicare	0.968	0.175	0	1
Medicaid	0.953	0.212	0	1
ResCouncil	0.984	0.125	0	1
FamCouncil	0.224	0.417	0	1

ESTIMATION RESULTS

We estimate equation (1) by different methods, as shown in Table 4. The first column shows the estimation results for the pooled data. We also take nursing homes' fixed effects into account and run a panel data regression, the estimates of which are shown in the second column. Since many variables are time-invariant, we cannot estimate their coefficients directly through the fixed-effect method but instead implement Hausman-Taylor method, as shown in the third column. In the estimates from all methods, the main effect, *Incentive,* is positive and statistically significant, which indicates that nursing homes with higher financial incentives are more likely to improve their star ratings after self-reporting.

Table 4. Estimates of equation (1)

Variables	Pooled data	Fixed effect	Hausman-Taylor
Incentive	0.0821 *** (0.0123)	0.102 *** (0.00996)	0.102 *** (0.00993)
Competition	0.000781 (0.000712)	-	-0.000122 (0.00963)
ResTotal	0.000798 (0.00064)	-	0.000205 (0.000804)
ForProfit	-0.175 ** (0.0881)	-	-0.107 (0.107)
Medicare	-1.28 *** (0.147)	-	-0.995 *** (0.187)
Medicaid	-0.373 *** (0.132)	-	-0.138 (0.19)
ResCouncil	-0.086 (0.192)	-	-0.0474 (0.213)
FamCouncil	-0.261 *** (0.0654)	-	-0.131 (0.116)

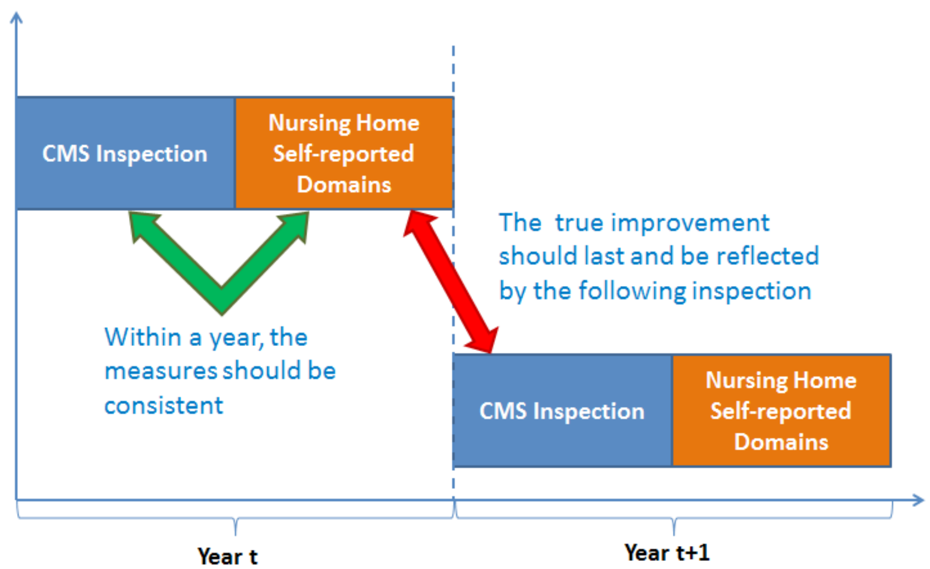
*Note: *: significant at p<0.1; **: significant at p<0.1; ***: significant at p<0.1*

INFLATION DETECTION AND DEMONSTRATION

CORRELATION ANALYSIS

Although the preliminary results show a positive association between financial incentive and changes in star-ratings, they do not necessarily indicate inflation in self-reported measures. It is possible that nursing homes gain the additional stars legitimately through their true efforts. To explore the underlying reasons for the changes in ratings, we investigate the correlation between the on-site inspections and self-reported domains. As illustrated in Figure 4, under the assumption that there is no inflation and nursing homes self-reported measures are legitimate, positive correlations are expected between two sets of ratings. First, within the same year, a positive correlation is expected between the star ratings from CMS on-site inspection and those of nursing homes' self-reported domains. Second, if a nursing home really puts an effort toward improving its care quality, these efforts should have a lasting effect and lead to better results in the next year's on-site inspections and thus there should be a positive correlation between the star ratings from self-reported domains in one year and on-site inspection ratings in the subsequent year.

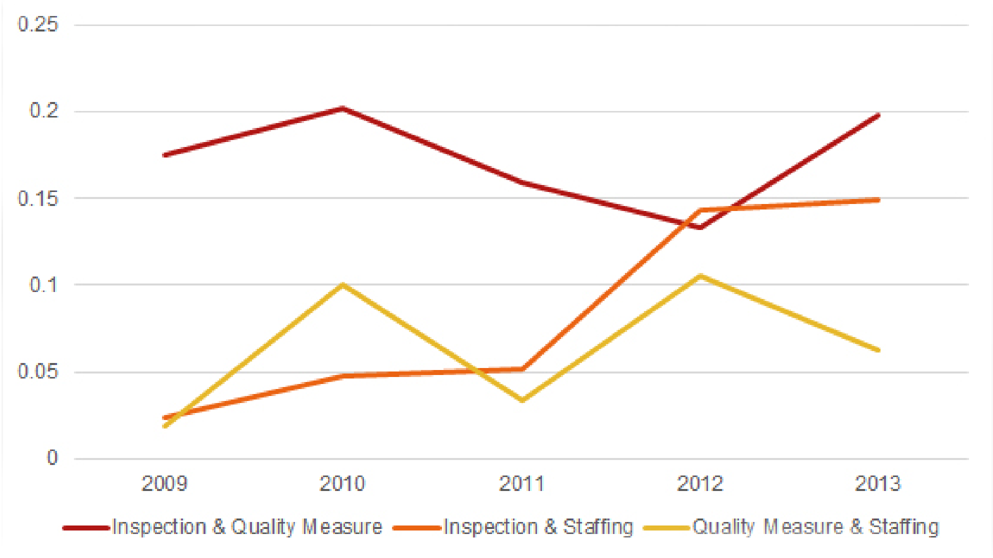
Figure 4. Graphical representation of correlation analysis



Note: If star increase is resulted from legitimate efforts, then a positive correlation is expected between self-reported measures in year 1 and on-site inspections in year 2 (red arrow). A positive correlation is also expected between the self-reported measures and the on-site inspection ratings in the same year (green arrow).

Figure 5 shows the correlations within the same year for the three pairs of domains over five consecutive years. It can be seen that all three pairs of correlations stay at a low level, around 0.2. The result clearly indicates inconsistency between the on-site inspections and self-reported domains within the same year. The self-reported measures and the next year's on-site inspection even shows a slightly negative correlation of -0.094, which indicates that the self-reported improvements in quality measures and staffing domains have no lasting effect on the next year's on-site inspection results at all. The correlation analysis serves as a preliminary evidence of potential inflation, and triggers our further analysis.

Figure 5. Correlation analysis for 5 consecutive years



Note: The red line represents the correlation between on-site inspection and quality measure. The orange line represents the correlation between on-site inspection and staffing. The yellow line represents the correlation between quality measure and staffing.

COMPLAINT-BASED ANALYSIS

In this section, we conduct further analysis to justify the existence of rating inflation. We identify a quantifiable third-party proxy variable which can serve as an independent measure of service quality, and compare the results with the star ratings given by the rating system. If significant inconsistency exists between the two, then the star ratings are questionable, and rating inflation likely exists. In our method, we use the number of complaints, which has been used as a common measure of the service quality in the literature of service and complaint management in many service industries (E. Anderson, Claes, and Roland 1997; Gardner 2004; Johnson 2001; Roland and Chung 2006). Specifically, we conduct an analysis based on the CDPH complaint data, which is an independently collected data set of patient complaints against California nursing homes.

If inflation does not exist, then the overall rating should be consistent with the true service quality, which is reflected by the number of complaints. That is, for nursing homes with the same overall rating, we expect them to have similar service qualities and a similar number of complaints. Table 5 shows the average number of complaints for nursing homes with different on-site inspection and overall ratings. For each overall rating level, the nursing homes are divided into two categories: Nursing homes whose star ratings increased after self-reporting and nursing homes whose star ratings did not increase after self-reporting. We denote the upper triangular section as area *I (shaded)* and lower triangular section as area *II*. The shaded area (*I*) includes those nursing homes whose overall rating has increased as a result of their self-reported measures. Area *II* includes those nursing homes whose overall rating either decreased or remained the same after self-reporting. This classification allows us to test the following claims:

Claim one: If the improvements observed are not resulted from legitimate efforts and inflation does exist, nursing homes with the same overall star rating but different on-site inspection ratings should have different complaint distributions.

Table 5. Average number of patient complaints

		Overall star rating				
		1	2	3	4	5
On-site inspection stars	1	7.981	6.989			
	2	6.193	6.271	6.010	8.389	
	3	3.929	3.934	4.633	4.940	4.056
	4		3.923	3.799	3.503	2.826
	5			6.667	2.157	2.423

Notes: [a]The blank cells represent the impossible rating transaction according to CMS's rating system design. [b] The shaded cells represent nursing homes of which the ratings increased after self-reporting. The inflators are among these nursing homes.

The results of two ANOVA tests are presented in Table 6. In the first column, nursing homes with the same overall ratings are grouped by whether or not their star rating increased after self-reporting. In other words, we examine if the shaded and unshaded cells in each column of Table 5 have similar distributions. In the second column, we group nursing homes with the same overall rating based on their on-site inspection ratings. In other words, we examine if all the cells in each column of Table 5 have similar distributions. As reported in Table 6, all the comparisons are significant and thus the claim that nursing homes with the same overall rating but different inspection ratings have different complaint distributions is supported.

Table 6. F statistics: Comparison in each overall rating

		Grouped by Area I vs Area II	Grouped by inspection ratings
Overall star rating	1	-	4.61**
	2	7.43***	6.16***
	3	13.05***	5.06***
	4	14.22***	8.35***
	5	5.27**	5.70***

*Note: *: significant at p<0.1; **: significant at p<0.1; ***: significant at p<0.1*

Claim two: If the improvements observed are not resulted from legitimate efforts and inflation does exist, nursing homes with the same inspection rating but different overall ratings should have similar complaint distributions.

The results of two ANOVA tests are presented in Table 7. In the first column, nursing homes with the same on-site inspection ratings are grouped by whether or not their star rating increased after self-reporting. In other words, we examine if the shaded and unshaded cells in each *row* of Table 5 have similar distributions. In the second column, we group nursing homes with the same inspection rating based on their overall star ratings. In other words, we examine if all the cells in each row of Table 5 have similar distributions. As shown in Table 7, we do not observe a significant difference in the number of complaints, although the overall rating can be quite different. The results show that service quality does not improve for nursing homes whose star ratings get improved after self-reporting and thus claim two is also supported. Together with the results obtained for claim one, the analysis provides strong evidence of the existence of rating inflation in self-reported measures.

Table 7. F statistics: Comparison in each inspection rating

		Grouped by Area I vs Area II	*Grouped by overall ratings*
Inspection ratings	1	2.46	2.46
	2	0.12	0.12
	3	0.78	0.78
	4	5.37**	2.00
	5	-	1.91

*Note: *: significant at p<0.1; **: significant at p<0.1; ***: significant at p<0.1*

PREDICTION MODEL AND VARIABLE IMPORTANCE ANALYSIS

In this section, we first develop a method which gives a quantifiable estimate of the extensiveness of rating inflation. We then run a variable importance analysis to summarize key characteristics of the likely inflators.

For a nursing home that inflates its self-reported measures, the overall rating is driven by two components. The first component is the observable characteristics which are common between cheating and honest nursing homes. The second component is the unobservable inflation coefficient which only pertains to the inflating nursing homes. If we model the overall ratings as a function of observed characteristics, the inflation component is unobserved and omitted from our regression model, thus the estimates of the remaining observed variables will suffer from the omitted variable bias. However, since the overall star ratings of honest nursing homes are only driven by one component of observed characteristics and the inflation component does not exist among the honest nursing homes, our regression estimates for the honest group will not suffer from the omitted variable bias. To develop our inflation prediction model, we first divide the nursing homes into two groups: the honest nursing homes and the remaining, defined as potential inflators. A regression is then run for the honest nursing homes. The obtained regression coefficients from the sample of honest nursing homes are unbiased and reflect the true associations without inflation. These unbiased coefficients are then used to predict the highest possible overall star rating for each nursing home in the suspected inflating group. A nursing home is identified as a likely inflator in our estimation if its actual overall rating is higher than the highest level of its predicted overall rating.

PREDICTION MODEL

In our model, the overall star rating is used as the dependent variable, denoted by *OverallRating*. Similar to the variable *StarChange* in the regression model in section two, *OverallRating* is ordinal and takes values in five levels so we employ an ordinal logistic regression model. *OverallRating* is determined by a set of parameters γ_1, γ_2, γ_3, γ_4, which define the cut points of the five star levels. The model can be written as

$$P(Overallrating \leq k) = \frac{\exp(\gamma_k + x' \beta_P)}{1 + \exp(\gamma_k + x' \beta_P)}, \quad (2)$$

where k ϵ {1,2,3,4}. The independent variables denoted by vector are the same as the ones used in equation (1). The coefficients of prediction model are denoted by $\boldsymbol{\beta}_P$.

Since we use the coefficients of the honest group as the unbiased baseline, we define the members in this group very strictly to guarantee that there is no evidence of inflation for all nursing homes in the honest group. An honest nursing home is selected based on the following criteria:

1. Its overall star rating does not increase after self-reporting
2. The number of its patient complaints is strictly lower than the median of its corresponding self-reporting level.

To define a nursing home's self-reporting level, we combine the two self-reported domains by taking the average star rating of staffing and quality measure. The median and average of the number of complaints for each self-reporting level are reported in Table 8. Based on the two criteria, the honest (*H*) group consists of 1,345 nursing home records in five years. The remaining 3,033 nursing home records are categorized in the potential inflator (*PI*) group. Note that the PI group consists of both the actual inflators and the nursing homes who improve their service qualities through legitimate efforts. In the following, we estimate the proportion of the actual inflators in the *PI* population.

Table 8. Combined self-reported measures

Average Rating in self-reported domains	*Patient complaints*	
	Average	Median
Level 1: 1-1.5	6.073	4
Level 2: 2-2.5	5.621	3
Level 3: 3-3.5	5.183	3
Level 4: 4-4.5	4.578	2
Level 5: 5	3.088	1.5

We run the ordinal logistic regression in equation (2) on the sample of honest nursing homes (H group) to obtain the unbiased estimates of each coefficient, as shown in Table 9. Both the 95e percent and 90 percent confidence intervals are reported. Using the upper bounds of unbiased coefficient estimates, we then predict the highest possible rating for each of the nursing homes in the *PI* group. A nursing home is classified as an inflator if its actual overall star rating is higher than the highest possible rating predicted through our model. Based on the 95 percent confidence interval, we can identify 184 inflator records out of the 3,033 nursing home records (6.07 percent) in the PI group. Based on the 90 percent confidence interval, we can identify 379 inflator records (12.5 percent) in the PI group.

Table 9. Unbiased coefficient estimates for the H Group

Variables	Coefficients	95% Confidence interval		90% Confidence interval	
Incentive	-0.234*** (0.026)	-0.285	-0.183	-0.277	-0.191
Competition	-0.002 (0.001)	-0.004	0.000626	-0.00391	0.000230
ResTotal	-0.139*** (0.001)	-0.168	-0.0111	-0.0163	-0.0115
ForProfit	-1.349*** (0.142)	-1.627	-1.072	-1.582	-1.117
Medicare	-1.784*** (0.399)	-2.565	-1.002	-2.439	-1.128
Medicaid	-0.470* (0.251)	-0.962	0.0214	-0.883	-0.0576
ResCouncil	-0.154 (0.379)	-0.896	0.588	-0.777	0.469
FamCouncil	0.483*** (0.115)	0.258	0.708	0.294	0.672

*Note: *: significant at p<0.1; **: significant at p<0.1; ***: significant at p<0.1*

VARIABLE IMPORTANCE ANALYSIS

It is important to understand the key differences between honest nursing homes and the inflators, so that we can focus on these differences in audits and identify the inflators efficiently. In this section, a variable importance analysis is conducted to explore the key characteristics of the inflators. A subset of the data is first constructed by eliminating nursing homes whose status cannot be identified. The eliminated nursing homes are the ones which are neither identified as inflators nor are in the honest group. The remaining dataset consists of 1,724 nursing home records, in which 1,345 records are for honest nursing homes and 379 records are for the likely inflators identified using a 90 percent confidence interval. The status of a nursing home is assigned as zero if it belongs to the honest group and one if it is a likely inflator. To perform the variable importance analysis, we use the logistic specification presented in equation (3).

$$logit(\lambda) = \boldsymbol{x}_{it}^{'}\boldsymbol{\beta} + \boldsymbol{y}_{it}^{'}\boldsymbol{\tau} \qquad (3)$$

where λ is the probability of being identified as an inflator, is the vector of variables that were also used in equations (1) and (2) and is a vector of ratings in the three domains. The coefficient estimates and their importance are presented in Table 10. Among the variables, we find the variable *ResTotal* to be the top in terms of variable importance. The result indicates that when a nursing home's size grows, its probability to game the rating system increases significantly. The *Incentive* and *ForProfit*, being the next two most important characteristics in the variable importance analysis, indicate that the financial incentive of a nursing home plays an important role in being an inflator. The for-profit nursing homes are more likely to inflate their self-reported ratings than the non-profits ones, and the higher their financial incentives are, the more likely they will be inflators. This is consistent with the work of Chesteen et al. (2005) which shows that non-profit nursing homes have a higher quality of service. The probability of being an inflator does not significantly change with competition and certification status.

Table 10. Variable Importance Analysis

Variables	*Coefficients*	*Variable Importance*
Incentive	0.0336*** (0.0033)	45.655
Competition	0.0000567 (0.00019)	0.000
ResTotal	0.00239*** (0.00016)	69.928
ForProfit	0.223*** (0.0225)	44.715
Medicare	0.0366 (0.0419)	2.674
Medicaid	0.0132 (0.0372)	0.271
ResCouncil	-0.142** (0.052)	11.294
FamCouncil	-0.141*** (0.0174)	36.318
InspectionRating	0.00182** (0.0695)	10.791
QualityRating	0.0123*** (0.0564)	100.000
StaffingRating	0.0125*** (0.0656)	87.057

*Note: *: significant at p<0.1; **: significant at p<0.1; ***: significant at p<0.1*

CONCLUSION

This paper systematically analyzes CMS's nursing home rating system, demonstrates the existence of inflation, and presents a model to detect likely inflators. We show that nursing homes have strong financial incentives directly related to higher star ratings, which may in turn drive the inflating behaviors. We then develop a systematical method which uses the independent third-party measure of patient complaints to demonstrate the existence of rating inflation. An inflation prediction model is then developed, which provides an estimate of the proportion of inflating nursing homes in the current system, and gives a quantifiable evaluation of the system performance. The variable importance analysis is then performed to identify the factors that contribute the most to being an inflator.

Our research provides several contributions. First, to the best of our knowledge, this is the first study that systematically investigates the inflation in the CMS nursing home rating system. It explores the fundamental financial reason for a nursing home to improve the star rating, even by inflating self-reported measures, which links the dots between incentives and observed behavior. Second, we contribute to the theory by developing a systematical method for demonstrating the existence of rating inflation and evaluating inflator proportion. As we discussed earlier, although CMS has implemented minor improvements to its rating system, it is still largely based on self-reported measures and does not address the issue of inflation. Our research demonstrates the shortcoming of the rating system and informs CMS on how to improve its system or how to identify the likely inflators. This study estimates the proportion of likely inflators and summarizes their key characteristics. The results can be used to strategically focus future audits on the nursing homes which are most likely to be inflators, and help CMS improve the rating system.

This work also has several limitations. First, we are unable to measure the financial incentives for each nursing home at the individual level. This is practically very difficult, since even for the same nursing home at the same rating level, the financial incentive may vary over time depending on various financial situations. To address this limitation, we perform our analysis on an aggregated level and use the average as a universal incentive for each rating group, leaving the unobserved incentive fluctuations to the nursing homes fixed effects. Second, we do not observe self-reporting inflation directly and can only infer it from an aggregated level. This is a common issue in misbehavior detection research due to the unavailability of individual-level data. We address this limitation by calculating the highest possible rating using the confidence interval and using the most conservative statistics. Third, we are only able to measure patient complaints in numbers, but not in “severeness.” For example, a complaint about medical malpractice may have much more impact than a complaint about sanitation. Future research can apply text mining techniques to address this limitation.

REFERENCES

Anderson, David, Guodong (Gordon) Gao, and Bruce Golden. 2014. "Life Is All about Timing: An Examination of Differences in Treatment Quality for Trauma Patients Based on Hospital Arrival Time." *Production and Operations Management* 23 (12): 2178–90. doi:10.1111/poms.12236.

Anderson, Eugene, Fornell Claes, and Rust Roland. 1997. "Customer Satisfaction, Productivity, and Profitability: Differences between Goods and Services." *Marketing Science* 16 (2): 129–45.

Andritsos, Dimitrios A., and Sam Aflaki. 2015. "Competition and The Operational Performance of Hospitals: The Role of Hospital Objectives." *Production and Operations Management* 24 (11): 1812–32. doi:10.1111/poms.12416.

Bauer, Thomas, and Sinning Mathias. 2008. "An Extension of the Blinder-Oaxaca Decomposition to Nonlinear Models." *AStA Advances in Statistical Analysis* 92 (2): 197–206.

Berg, Katherine, Vincent Mor, John Morris, Katharine Murphy, Terry Moore, and Yael Harris. 2002. "Identification and Evaluation of Existing Nursing Homes Quality Indicators." *Healthcare Financing Review* 23.4.

Cayirli, Tugba, and Emre Veral. 2003. "Outpatient Scheduling in Health Care: A Review of Literature." *Production and Operations Management* 12 (4): 519–49. doi:10.1111/j.1937-5956.2003.tb00218.x.

Charlene, Harrington, Steffie Woolhandler, Joseph Mullan, Helen Carrillo, and David Himmelstein. 2001. "Does Investor Ownership of Nursing Homes Compromise the Quality of Care?" *American Journal of Public Health* 91 (9): 1452–55.

Chen, Hong, Qu Qian, and Anming Zhang. 2015. "Would Allowing Privately Funded Health Care Reduce Public Waiting Time? Theory and Empirical Evidence from Canadian Joint Replacement Surgery Data." *Production and Operations Management* 24 (4): 605–18. doi:10.1111/poms.12260.

Chesteen, Susan, Berit Helgheim, Taylor Randall, and Don Wardell. 2005. "Comparing Quality of Care in Non-Profit and for-Profit Nursing Homes: A Process Perspective." *Journal of Operations Management*, Operations Management in Not-For-Profit, Public and Government Services, 23 (2): 229–42. doi:10.1016/j.jom.2004.08.004.

Chow, Vincent S., Martin L. Puterman, Neda Salehirad, Wenhai Huang, and Derek Atkins. 2011. "Reducing Surgical Ward Congestion Through Improved Surgical Scheduling and Uncapacitated Simulation." *Production and Operations Management* 20 (3): 418–30. doi:10.1111/j.1937-5956.2011.01226.x.

Committee on Nursing Home Regulation. 1986. *Improving the Quality of Care in Nursing Homes*. Institute of Medicine.

David, Grabowski, Joseph Angelelli, and Vincent Mor. 2004. "Medicaid Payment and Risk-Adjusted Nursing Home Quality Measures." *Health Affairs* 23 (5): 243–52.

David, Grabowski, Zhanlian Feng, Orna Intrator, and Vincent Mor. 2004. "Recent Trends in State Nursing Home Payment Policies." *Health Affairs* w4.

DellaVigna, Stefano, and Eliana La Ferrara. 2010. "Detecting Illegal Arms Trade." *American Economic Journal: Economic Policy* 2 (4): 26–57.

Dobson, Gregory, Sameer Hasija, and Edieal J. Pinker. 2011. "Reserving Capacity for Urgent Patients in Primary Care." *Production and Operations Management* 20 (3): 456–73. doi:10.1111/j.1937-5956.2011.01227.x.

Duggan, Mark, and Levitt Steven. 2000. "Winning Isn't Everything: Corruption in Sumo Wrestling." *National Bureau of Economic Research* w7798.

Duhigg, Charles. 2007. "Washington Scrutinizes Nursing Homes." *The New York Times*, November 16. http://www.nytimes.com/2007/11/16/business/16care.html.

Eiken, Steve, Kate Sredl, Lisa Gold, Jessica Kasten, Brian Burwell, and Paul Saucier. 2014. "Medicaid Expenditures for Long-Term Services and Supports in FFY 2012." *Bethesda, MD: Truven Health Analytics*. http://www1a.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Long-Term-Services-and-Supports/Downloads/LTSS-Expenditures-2012.pdf.

Engelberg, Joseph, Parsons Christopher, and Tefft Nathan. 2014. "Financial Conflicts of Interest in Medicine." *SSRN* 2297094.

Fairlie, Robert. 2005. "An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models." *Journal of Economic and Social Measurement* 30 (4): 305–16.

Fennell, Mary, Zhanlian Feng, Melissa Clark, and Vincent Mor. 2010. "Elderly Hispanics More Likely to Reside in Poor-Quality Nursing Homes." *Health Affairs* 29 (1): 65–73.

Fowles, Donald G. 2012. "A Profile of Older Americans." *Administration on Aging*, US Department of Health and Human Services.

Gardner, Everette. 2004. "Dimensional Analysis of Airline Quality." *Interfaces* 34 (4): 272–79.

Güneş, Evrim Didem, Egemen Lerzan Örmeci, and Derya Kunduzcu. 2015. "Preventing and Diagnosing Colorectal Cancer with a Limited Colonoscopy Resource." *Production and Operations Management* 24 (1): 1–20. doi:10.1111/poms.12206.

Helm, Jonathan E., Shervin AhmadBeygi, and Mark P. Van Oyen. 2011. "Design and Analysis of Hospital Admission Control for Operational Effectiveness." *Production and Operations Management* 20 (3): 359–74. doi:10.1111/j.1937-5956.2011.01231.x.

Jack, Eric P., and Thomas L. Powers. 2004. "Volume Flexible Strategies in Health Services: A Research Framework." *Production and Operations Management* 13 (3): 230–44. doi:10.1111/j.1937-5956.2004.tb00508.x.

Jacob, Brian, and Levitt Steven. 2003. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *National Bureau of Economic Research* w9413.

Johnson, Robert. 2001. "Linking Complaint Management to Profit." *International Journal of Service Industry Management* 12 (1): 60–69.

LaGanga, Linda R., and Stephen R. Lawrence. 2012. "Appointment Overbooking in Health Care Clinics to Improve Patient Service and Clinic Performance." *Production and Operations Management* 21 (5): 874–88. doi:10.1111/j.1937-5956.2011.01308.x.

Mayzlin, Dina, Dover Yaniv, and Chevalier Judith. 2012. "Promotional Reviews: An Empirical Investigation of Online Review Manipulation." *National Bureau of Economic Research* w18340.

McCoy, Jessica H., and M. Eric Johnson. 2014. "Clinic Capacity Management: Planning Treatment Programs That Incorporate Adherence." *Production and Operations Management* 23 (1): 1–18. doi:10.1111/poms.12036.

Medicare, Centers for, Medicaid Services 7500 Security Boulevard Baltimore, and Md21244 Usa. 2016. "Staffing-Data-Submission-PBJ." March 17. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/Staffing-Data-Submission-PBJ.html.

Mor, Vincent, Katherine Berg, Joseph Angelelli, David Gifford, John Morris, and Terry Moore. 2003. "The Quality of Quality Measurement in US Nursing Homes." *The Gerontologist* 43 (suppl 2): 37–46.

National Center for Health Statistics (U.S.), ed. 2009. *The National Nursing Home Survey: 2004 Overview*. Vital and Health Statistics. Ser. 13, Data from the National Health Care Survey, no. 167. Hyattsville, Md: U.S. Dept. of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review*, 693–709.

OBRA. 1987. "Federal Nursing Home Reform Act from the Omnibus Budget Reconciliation Act of 1987." US Congress.

Roland, Rust, and T. Siong Chung. 2006. "Marketing Models of Service and Relationships." *Marketing Science* 25 (6): 560–80.

Rosalie, Kane. 2003. "Definition, Measurement, and Correlates of Quality of Life in Nursing Homes: Toward A Reasonable Practice, Research, and Policy Agenda." *The Gerontologist* 43 (suppl 2): 28–36.

Roth, Aleda V., and Roland Van Dierdonck. 1995. "Hospital Resource Planning: Concepts, Feasibility, and Framework." *Production and Operations Management* 4 (1): 2–29. doi:10.1111/j.1937-5956.1995.tb00038.x.

Salzarulo, Peter A., Stephen Mahar, and Sachin Modi. 2015. "Beyond Patient Classification: Using Individual Patient Characteristics in Appointment Scheduling." *Production and Operations Management*, December, n/a – n/a. doi:10.1111/poms.12528.

Smith, David Barton, Zhanlian Feng, Mary Fennell, Jacqueline Zinn, and Vincent Mor. 2007. "Separate and Unequal: Racial Segregation and Disparities in Quality Across US Nursing Homes." *Health Affairs* 26 (5): 1448–58.

Stevenson, David, and Grabowski David. 2008. "Private Equity Investment and Nursing Home Care: Is It A Big Deal?" *Health Affairs* 27 (5): 1399–1408.

Stevenson, David, and Studdert David. 2003. "The Rise of Nursing Home Litigation: Findings from A National Survey of Attorneys." *Health Affairs* 22 (2): 219–29.

Thomas, Katie. 2014. "Medicare Star Ratings Allow Nursing Homes to Game the System." *The New York Times*, August 24. http://www.nytimes.com/2014/08/25/business/medicare-star-ratings-allow-nursing-homes-to-game-the-system.html.

Werner, Rachel, and Tamara Konetzka. 2010. "Advancing Nursing Home Quality Through Quality Improvement Itself." *Health Affairs* 29 (January): 81–86.

Werner, Rachel M., R. Tamara Konetzka, and Daniel Polsky. 2016. "Changes in Consumer Demand Following Public Reporting of Summary Quality Ratings: An Evaluation in Nursing Homes." *Health Services Research*, February. doi:10.1111/1475-6773.12459.

Werner, Rachel, Elizabeth Stuart, and Daniel Polsky. 2010. "Public Reporting Drove Quality Gains at Nursing Homes." *Health Affairs* 29 (9): 1706–13.

William, Cabin, David Himmelstein, Michael Siman, and Steffie Woolhandler. 2014. "For-Profit Medicare Home Health Agencies' Costs Appear Higher and Quality Appears Lower Compared to Nonprofit Agencies." *Health Affairs* 33 (8): 1460–65.

Zepeda, E. David, and Kingshuk K. Sinha. 2016. "Toward an Effective Design of Behavioral Health Care Delivery: An Empirical Analysis of Care for Depression." *Production and Operations Management*, January, n/a – n/a. doi:10.1111/poms.12529.

Zhang, Yue, Martin L. Puterman, Matthew Nelson, and Derek Atkins. 2012. "A Simulation Optimization Approach to Long-Term Care Capacity Planning." *Oper. Res.* 60 (2): 249–61. doi:10.1287/opre.1110.1026.

ENDNOTES

1 The annual on-site inspection looks into areas such as medication management, nursing home administration, environment, food service, and residents' rights and quality of life.

2 The Staffing domain is evaluated based on the self-reported CMS Certification and Survey Provider Enhanced Reports (CASPER) staffing data. The two measures used are the total nursing hours and Registered Nursing (RN) hours, and are adjusted for case-mix based on the Resource Utility Group (RUG-III) case-mix system derived from the Minimum Data Set (MDS). The staffing star rating is then updated by the end of the quarter when raw data is collected.

3 The Quality Measure domain rating uses nine out of 18 quality measurement criteria developed from the MDS, which covers seven aspects from long-stay terms and 2 aspects from short-stay terms. The quality measure star rating is then updated by the end of each quarter by using the results from three most recent quarters.

4 Both four and five stars in staffing rating are qualified for obtaining additional overall star rating, while only five stars in quality measure is qualified. Additional conditions apply to nursing homes whose inspection ratings are only one star, and for nursing homes which are in the CMS's Special Focus Facility (SFF) program. The overall star rating cannot be more than five stars or less than one star.

5 According to CMS's rating mechanism design, the top 10 percent nursing homes in inspection receive five stars. The bottom 20 percent nursing homes in inspection receive one star. Nursing homes which rank in between receive two-to-four stars according to a fixed proportion. There is no fixed proportion for the self-reported domains.

6 CMS does not consider star ratings in its Nursing Facility Prospective Payment System (PPS) which determines Medicare's payment amount for a particular service (CMS, Prospective Payment Systems - General Information, 2015) (CMS, Skilled Nursing Facility Prospective Payment System, 1997). However, nursing homes' revenue from non-Medicare patients (self-paying and other resources) and other non-health care services (e.g., accommodation and dining) can significantly affect their overall profits. Moreover, the increased demand for their services that happens as a result of high star ratings (R. M. Werner, Konetzka, and Polsky 2016) can reduce their overhead costs and thus lead to an increase in their per-patient profit.

EMAIL YOUR COMMENTS TO GSCOMMENTS@BROOKINGS.EDU

GOVERNANCE STUDIES

The Brookings Institution
1775 Massachusetts Ave., NW
Washington, DC 20036
Tel: 202.797.6090
Fax: 202.797.6144
brookings.edu/governance.

EDITING
Liz Sablich

PRODUCTION & LAYOUT
Cathy Howell

Support for this publication was generously provided by the National Institute for Health Care Management (NIHCM) Research and Education Foundation.