POMS PRODUCTION AND OPERATIONS MANAGEMENT

PRODUCTION AND OPERATIONS MANAGEMENT POMS

Vol. 0, No. 0, xxxx–xxxx 2017, pp. 1–19
ISSN 1059-1478 | EISSN 1937-5956 | 17 | 00 | 0001
DOI 10.1111/poms.12804

Winning at All Costs: Analysis of Inflation in Nursing Homes' Rating System

Xu Han
Department of Information Technology & Operations Management, College of Business, Florida Atlantic University, 777 Glades Rd, Boca Raton, FL 33431, USA, hanx@fau.edu

Niam Yaraghi*
Operations and Information Management Department, School of Business, University of Connecticut, 2100 Hillside Rd, Storrs, CT 06268, USA
Center for Technology Innovation, Governance Studies, The Brookings Institution, 1775 Massachusetts Ave NW, Washington, DC 20036, USA, niam.yaraghi@uconn.edu

Ram Gopal
Operations and Information Management Department, School of Business, University of Connecticut, 2100 Hillside Rd, Storrs, CT 06268, USA, rgopal@business.uconn.edu

The Nursing Home Compare system administrated by the Centers for Medicare & Medicaid Services (CMS) is widely used by patients, medical providers and payers. We argue that the rating system is prone to inflation in self-reported measures, which leads to biased and misleading ratings. We use the CMS rating data over 2009–2013 and the corresponding financial data reported by Office of Statewide Health Planning and Development and patients' complaints data reported by California Department of Public Health for 1219 nursing homes in California to empirically examine the key factors affecting the star rating of a nursing home. We find a significant association between the changes in a nursing home's star rating and its profits, which points to a financial incentive for nursing homes to improve the ratings. We then demonstrate that this association does not always lead to legitimate efforts to improve service quality, but instead can induce inflation in self-reporting in the rating procedure. A prediction model is then developed to evaluate the extensiveness of inflation among the suspect population based on which 6% to 8.5% of the nursing homes are identified as likely inflators. We also summarize the key characteristics of likely inflators, which can be useful for future audit.

Key words: inflation detection; nursing home; rating system
History: Received: March 2016; Accepted: September 2017 by Subodha Kumar, after 1 revision.

1. Introduction

Nearly two million Americans spend an average of 835 days of their life in one of the 15,700 nursing home facilities in the United States (Bercovitz et al. 2009). The Department of Health and Human Services estimates that in 2009, 4.1% of Americans over 65 years old lived in these facilities. This percentage increases with age, ranging from 1.1% in the population of 65 to 74 years old to 13.2% 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 where 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 publicly 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 1-star facilities decreased by 8%, while the market share of 5-star facilities increased by more than 6%.

Given the important role of CMS's ratings, nursing homes have a significant incentive to improve their ratings. However, these ratings may not always reflect true quality. Cases have been reported in which highly rated nursing homes only provide substandard care, which even lead to the death of patients.

Please Cite this article in press as: Han, X., et al. Winning at All Costs: Analysis of Inflation in Nursing Homes' Rating System. *Production and Operations Management* (2017), https://doi.org/10.1111/poms.12804

This research is based on the publicly available data provided by multiple government agencies including CMS, California Statewide Health Planning and Development (OSHPD), and 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 CMS inspection rating and self-reported ratings within the same year to be closely associated. Our correlation analysis results, however, shows almost no correlation between the inspection and self-reported results, and sheds doubt on the legitimacy of self-reported measures. We 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. The result serves as strong evidence on the existence of inflation in the current rating system as it points to the fact that the service quality is predicted by the *Health 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 estimate the proportion of nursing homes that have inflated their self-reported ratings. Using a 95% confidence interval, we identify around 6% of nursing homes in the suspect population to be likely inflators in the current system. Fourth, we conduct a variable importance analysis to identify the factors whose change contributes the most to the probability of being an inflator. Our results demonstrate the shortcomings of the current rating systems and call for significant reforms in how CMS and other payers evaluate the quality of nursing homes.

The study proceeds as follows. In section 2, we discuss the background and conceptual framework of our research, including how the rating measures are established, how the star ratings are generated, and what the potential issues are with this rating mechanism. In this section, we also introduce a theoretical framework to address these issues. In section 3, we review related literature on nursing home quality measures, misbehavior detection, and quantifying methods. In section 4, we describe our data collection procedure and explore the underlying financial incentives for nursing homes to improve their ratings. In section 5, 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. We develop a prediction model in section 6 to identify likely rating inflators and evaluate the performance of the system. We then conduct a variable importance analysis to show key characteristics of the inflators. We conclude the whole study in section 7, and discuss the limitations and future work.

2. Background and Theoretical Framework

2.1. The History of Nursing Home Rating System

The standardization of nursing home service quality started even before CMS was founded. In 1961, the Public Health Service (PHS) began studying nursing homes' state licensures, after a series of problems being reported by the Commission on Chronic Illness from several states. The Nursing Home Standards Guide, issued by the PHS, specified 77 service standards in health and safety, which established the foundation of nursing home service standards. Since then, Nursing Home Standards Guide gradually developed and more standards were included. By 1974, a total of 90 standards were included, covering various aspects in health and safety. In 1977, the Health Care Financing Administration (HCFA) was created as a new federal organization, and continued the standardization and certification of nursing home service qualities. The HCFA commissioned the Institute of Medicine to examine the standards in nursing home services.

A major reform on nursing homes' regulation took place in 1987, when the Nursing Home Reform Act also known as OBRA-87 was passed (Turnham 2001). The OBRA-87 established more stringent inspection, and further specified and revised the regulations on nursing home services, including nurse training, care standards, sanctions, and remedies. It also established the use of Resident Assessment Instrument, of which the Minimum Data Set is a major component, and is widely used today in nursing home research.

Please Cite this article in press as: Han, X., et al. Winning at All Costs: Analysis of Inflation in Nursing Homes' Rating System. *Production and Operations Management* (2017), https://doi.org/10.1111/poms.12804

The HCFA changed its name to the CMS in 2001. CMS released its Nursing Home Compare (NHC) system in October 1998, in the form of report card, which provided information on Medicare/Medicaid-certified nursing homes via Internet. The initial system only included nursing homes' basic information and the deficiencies on health and safety found in inspection, which are also covered in today's *Health Inspection*. The *Staffing* measure was included in the system in June 2000, and the *Quality Measures* were included in November 2002 (General Accounting Office, 2002). This was the early form of today's 3-measure nursing home rating system. The NHC report card system was very influential since it was one of the earliest systems presenting publicly available standardized quality information on nursing homes. However, the report card method suffered challenges such as the lack of consumer awareness and access (Stevenson 2006), and the difficulties for consumers to understand the information on the report card (Shugarman and Brown 2006).

In order to address these issues, CMS launched its NHC system in December 2008, which is the current system being used. This reformed rating system followed the 3-measure setting in the previous report card system, but uses a 5-star scale on each of the three measures, which greatly improved the usability of the rating system. The 5-star nursing home rating system gradually became the gold standard of nursing home selection. As reported by CMS (2017a), the system gets more than 1.4 million visitors per year, with 85% users reporting that they found the information they are looking for on nursing homes. The dataset used in our research covers nursing homes' rating, complaint, and financial data from 2009 to 2013, which are the first 5 years since the inception of the 5-star rating system.

Centers for Medicare & Medicaid Services announced new policies to improve its nursing homes rating system starting in February 2015. These policies include expanding the targeted surveys, adding two additional measures in the *Quality Measures* domain, revising the staffing algorithm, and reporting payroll-based staffing information by the end of 2016 (CMS 2015). However, the framework of the 5-star rating system and the self-reporting process will not change, and therefore the conclusions of this research will remain relevant.

2.2. The Current Rating Mechanism

The CMS rating system is based on three domains: *Health Inspection*, *Staffing*, and *Quality Measures.* While independent, CMS-certified inspectors conduct and report the *Health Inspection*, 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 *Health Inspection* results. The *Health Inspection* looks into areas such as medication management, nursing home administration, environment, food service, and residents' rights and quality of life. Ratings are given based on the number, scope, and severity of deficiencies identified during the three most recent annual inspections (1/2 for current year, 1/3 for the previous year, and 1/6 for the second prior year). According to CMS's rating mechanism design, the top 10% nursing homes in *Health Inspection* receive 5 stars, while the bottom 20% nursing homes receive 1 star. Nursing homes which rank in between receive 2–4 stars in fixed proportions. There is no such restriction for self-reported measures.

Nursing homes are then assigned star ratings for the *Staffing* and *Quality Measures* domains. 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 Registered Nursing (RN) hours per resident day, and the total nursing hours, which is the sum of Registered Nurse (RN) hours, Licensed Practical Nurses (LPN) hours, and nurse aide hours per resident day. The results 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 are collected. The *Quality Measures* domain rating uses 9 out of 18 quality measurement criteria developed from the MDS, which covers seven aspects from long-stay terms and two aspects from short-stay terms. The *Quality Measures* data are also self-reported and is collected by the end of each quarter and the *Quality Measures* star rating is updated using the results from three most recent quarters.

The *Overall* star rating is then calculated by considering the *Health Inspection* rating as the baseline, adding 1 star if any self-reported domain is 5 stars and subtracting 1 star if any self-reported domain is 1 star. Nursing homes which only got 1 star in the *Health Inspection* can only have one additional star after self-reporting.[1] The *Overall* star rating cannot be more than 5 stars or less than 1 star. 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.

The detailed items covered in each measure are listed in Table 2. They evaluate nursing homes from three different angles. Generally speaking, the measures covered under *Health Inspection* reflect how organized the nursing facility is operating; the *Staffing* measures cover the number of working professionals in the facility, and the *Quality Measures* reflect how healthy the patients are living in the facility. Although measuring from different perspectives, close

Please Cite this article in press as: Han, X., et al. Winning at All Costs: Analysis of Inflation in Nursing Homes' Rating System. *Production and Operations Management* (2017), https://doi.org/10.1111/poms.12804

connections exist among these three domains (Castle 2008, Harrington et al. 2000, 2012, Kim et al. 2009, Konetzka et al. 2004, Munroe 1990, Zhang and Grabowski 2004). For example, urinary tract infection (UTI) is a common health problem found among nursing home patients, and the percentage of UTI is an important measure under the *Quality Measures* domain. Research has shown that UTI is closely related to catheter insertion (Gokula et al. 2004), which requires frequent and timely care, and as a result, an adequate level of staffing coverage. It has also been shown that the improper use of antibiotic agent is one of the major reasons causing UTI, and the antibiotic agent misusage is covered in the pharmacy service deficiencies, which are under the *Health Inspection* domain. As a result, UTI-associated problems are reflected in all the three domains, and it is often not possible to attribute such a complaint to any deficiencies in a particular domain. Similar examples can be found for other metrics such as pressure ulcer-associated problems. As a result, we argue that the three measures, although measuring from different angles, should be correlated at certain level. An unexpected low correlation is suspicious, and can be interpreted as a preliminary evidence of misreporting.

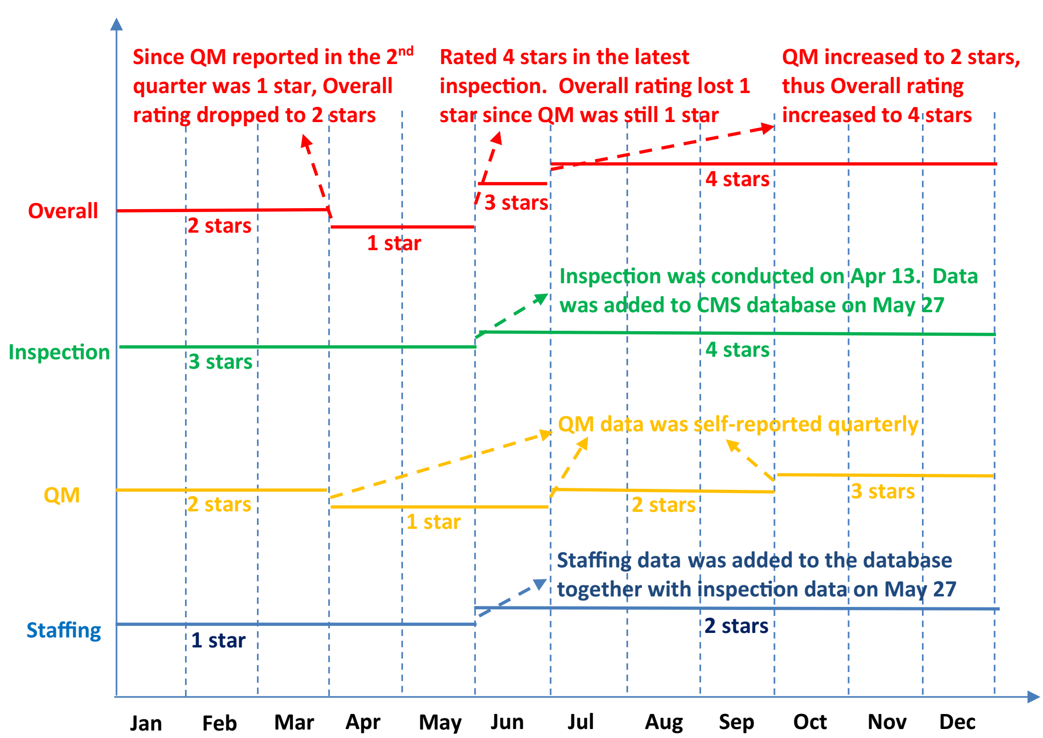
Table 1 An Example of a Nursing Home's Rating Dynamics

Month	Overall	Health inspection	Quality measures	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

2.3. Shortcomings of the Current Rating Mechanism

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 3 stars in the *Health Inspection* to gain two additional stars based on self-reported measures and

Figure 1 The Graphical Representation of a Nursing Home's Rating Dynamics [Color figure can be viewed at wileyonlinelibrary.com]



Note: [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 Measures 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.

Please Cite this article in press as: Han, X., et al. Winning at All Costs: Analysis of Inflation in Nursing Homes' Rating System. *Production and Operations Management* (2017), https://doi.org/10.1111/poms.12804

Table 2 Coverage of Each Measure (Health Inspection, Staffing, and Quality Measures)

Health Inspection (H: Health; F: Fire Safety)	Staffing	Quality Measures (L: long-stay; S: Short-stay)
Count of Administration Deficiencies (H)	RN hours/day	Percent of residents whose need for help with activities of daily living has increased (L)
Count of Environmental Deficiencies (H)	LPN hours/day	Percent of high-risk residents with pressure sores (L)
Count of Mistreatment Deficiencies (H)	Nurse aide hours/day	Percent of residents who have/had a catheter inserted and left in their bladder (L)
Count of Nutrition and Dietary Deficiencies (H)	Total Licensed hours/day	Percent of residents who were physically restrained (L)
Count of Pharmacy Service Deficiencies (H)	Total Nurse hours/day	Percent of residents with a urinary tract infection (L)
Count of Quality of Care Deficiencies (H)		Percent of residents who self-report moderate to severe pain (L)
Count of Resident Assessment Deficiencies (H)		Percent of residents experiencing one or more falls with major injury (L)
Count of Resident Rights Deficiencies (H)		Percent of residents with pressure ulcers that are new or worsened (S)
Count of Building Construction Deficiencies (F)		Percent of residents who self-report moderate to severe pain (S)
Count of Corridor Walls and Doors Deficiencies (F)		
Count of Electrical Deficiencies (F)		
Count of Emergency Plans and Fire Drills Deficiencies (F)		
Count of Exits and Egress Deficiencies (F)		
Count of Exit and Exit Access Deficiencies (F)		
Count of Fire Alarm Systems Deficiencies (F)		
Count of Furnishings and Decorations Deficiencies (F)		
Count of Hazardous Area Deficiencies (F)		
Count of Illumination and Emergency Power Deficiencies (F)		
Count of Interior Finish Deficiencies (F)		
Count of Laboratories Deficiencies (F)		
Count of Medical Gases and Anesthetizing Areas Deficiencies (F)		
Count of Miscellaneous Deficiencies (F)		
Count of Building Service Equipment Deficiencies (F)		
Count of Smoke Compartmentation and Control Deficiencies (F)		
Count of Smoking Regulations Deficiencies (F)		
Count of Automatic Sprinkler Systems Deficiencies (F)		
Count of Vertical Openings Deficiencies (F)		

become an excellent 5-star nursing home. As a result, the *Overall* rating can be quite different from the *Health Inspection* rating. Figure 2 shows how the ratings in each of these measures have shifted to higher stars during a period of 5 years from 2009 to 2013. By design, the proportion of *Health Inspection* star rating remains fixed in the 5 years, as shown in Figure 2a. However, the number of nursing homes that claim high performance in the self-reported domains has continuously increased over the past 5 years. As shown in Figure 2b, in 2009, about 40% of nursing homes self-reported to be 4 or 5 stars in the *Quality Measures* domain. This percentage has increased to 60% in 2013. On the other hand, about 20% of nursing homes self-reported to be 1 star in 2009, but less than 10% of nursing homes self-reported to be 1 star in 2013. In the *Staffing* domain, the number of highly rated nursing homes also significantly increased over this period, as shown in Figure 2c. Consequently, the *Overall* rating is consistently skewed to the higher end over time. As shown in Figure 2d, the portion of four or five nursing homes increased from 35% to 55% over the 5 years.

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, 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' 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 announced new policies to improve nursing homes' rating system (CMS 2015). These policies include the expansion of targeted surveys, including additional measures in the *Quality Measures* domain, and adding the payroll

Please Cite this article in press as: Han, X., et al. Winning at All Costs: Analysis of Inflation in Nursing Homes' Rating System. *Production and Operations Management* (2017), https://doi.org/10.1111/poms.12804

Figure 2 Distribution of Nursing Home Ratings from 2009 to 2013* [Color figure can be viewed at wileyonlinelibrary.com]

(a) *Health Inspection*

(b) *Quality Measures*

(c) *Staffing*

(d) *Overall*

Note: Colors represent different star rating groups.

information to the staffing reports. Despite these amendments, the structure of the rating system has not been changed and it still heavily relies on the self-reported domains and thus the newly revised system continues to be prone to manipulation by false self-reported measures. Whether the 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 is not clear. The objective of this research is to answer this question by investigating the existence and the extent of inflation in this rating system.

2.4. Proposed Methodology and Theoretical Framework

Data sources used by rating systems to generate ratings can be categorized into three groups: authority inspection (e.g., vehicle safety ratings), customer reporting (e.g., Amazon ratings), and self-reporting (e.g., business school rankings). CMS rates nursing homes by combining authority inspection and self-reported data in a unique way. To examine the reliability of such ratings, we use the third source of data (customer reporting) which was ignored in the original CMS ratings. In our rating inflation detection method, we use the number of patient complaints as a proxy of the true service quality (Carman 1990, Dabholkar et al. 1995, Tsaur et al. 2002). Under the assumptions that the self-reported measures are not inflated and reflect the honest efforts of nursing homes to improve their services, we expect the following:

1. If the ratings are not inflated, then for nursing homes with similar *Overall* ratings, we should expect similar service qualities, reflected in a similar number of complaints.

2. If the ratings are not inflated, we expect to observe an increase in the service quality of the nursing homes whose star ratings increased after self-reporting, comparing with nursing homes that initially had the same inspection rating but their ratings did not increase after self-reporting.

Our results, however, do not support any of the above inferences. We observe a clear difference in the number of complaints for nursing homes with the same *Overall* rating, indicating that their service qualities are quite different. We also observe no significant difference for nursing homes whose star rating increased after self-reporting, indicating that their reported improvements are highly questionable. The combined results indicate that the self-reported star rating increase cannot be simply explained by legitimate efforts, and rating inflation does exist.

To give a quantifiable estimate of the inflation, we incorporate ideas from the decomposition model developed by Oaxaca (Bauer and Mathias 2008, Fairlie 2005, Oaxaca 1973), which has been commonly used for quantifying group differences. Specifically, it decomposes the total difference between the groups into two parts: the differences caused by the differences in individual characteristics, and the differences caused by inconsistency in the measures. The model we developed is in line with the Oaxaca's idea. Using very restrictive criteria, which we will discuss later in the study, we divided the nursing homes into honest ones and potential inflators. We obtain the unbiased coefficient of the honest nursing homes and use the coefficients to predict the star ratings of the potential inflators. By doing this, we systematically control for differences caused by individual characteristics. A maximum predicted rating is then calculated for each of the potential inflators using selected confidence intervals, and is compared with the observed rating. If the observed rating is higher than the maximum predicted rating, then significant inconsistency exists in the measures, which points to the inflation of self-reported measures. By running this prediction model, we can identify likely inflators in the system and quantify the extent to which they have inflated their self-reported measures which in turn allow us to quantify the performance of the CMS' rating system.

3. Literature Review

The rating inflation problem is an important topic in many inter-connected fields, including healthcare facility operations management, healthcare policy research, and misbehavior detection. In this section, we first review literature in each related field, and then discuss the contribution of our work to the existing literature.

3.1. Healthcare Facility Operations Management

The research on the operations management of healthcare facilities includes an abundance of scholarly work, and research topics can be categorized based on their focused settings. The first stream of research analyzes the efficiency of operations and quality of care at hospitals. This line of research includes improving the patient scheduling systems (Cayirli and Veral 2003, Helm et al. 2011) and developing strategies to address the demand fluctuations (Jack and Powers 2004), analyzing the effects of patients' arrival time (Anderson et al. 2014) or the hospitals' objectives (Andritsos and Aflaki 2015) on the quality of care and creating alternative operations planning and control systems for curbing the increasing costs of hospital services (Roth and Van Dierdonck 1995). The other line of operations research focuses on individual physicians and small clinics. This stream includes design of public policies and novel scheduling strategies to reduce waiting time (Chen et al. 2015) and increase clinic performance (LaGanga and Lawrence 2012, Salzarulo et al. 2015), optimization of capacity and resources allocations and the effect of such improvements on quality of medical services in both primary (Dobson et al. 2011, McCoy and Eric Johnson 2014, Zepeda and Sinha 2016) and specialty care settings (Chow et al. 2011, Güneş et al. 2015).

Although nursing homes are an important part of the US health care system, operations and production management literature often neglects them. To the best of our knowledge, this line of research is limited to a few studies on minimizing waiting times (Zhang et al. 2012) and analyzing the effects of non-profit status of nursing homes on their service quality (Chesteen et al. 2005).

3.2. Healthcare Policy Research

The health policy literature related to nursing homes is rich, and we summarize them into three major categories. The first category of studies aims to examine how service quality can be quantifiably measured. Berg et al. (2002) evaluated existing quality indicators for long-term cares. Mor et al. (2003) used the MDS to point out that the incident-based nursing home quality measures can be unreliable. Shwartz et al. (2015) discussed the importance of using composite indexes to measure providers' performance. The second category of research mainly focuses on how to improve nursing home service qualities. Kane and Kane (2001) compared the senior patients and patients with disabilities to identify the key needs of senior residents in long-term cares. Walshe (2001) discussed how improper regulation can potentially detract from its effectiveness and lead to disappointing results. Grabowski and O'Malley (2014) discussed how

Please Cite this article in press as: Han, X., et al. Winning at All Costs: Analysis of Inflation in Nursing Homes' Rating System. *Production and Operations Management* (2017), https://doi.org/10.1111/poms.12804

telemedicine can reduce hospitalizations for nursing home residents. Stavropoulou et al. (2015) examined the performance of incident-reporting systems in improving patient safety. Many other studies, including the one by Mor et al. (2010) pointed out that the CMS payment incentives do not align incentives of care providers and care beneficiaries (Rosalie 2003, Mor et al. 2004, Werner and Konetzka 2010, Werner et al. 2010). The third category of research discusses problems existing in the current nursing home market, some of which are major barriers for enhancing service quality. These problems include racial segregation in nursing homes (Smith et al. 2007), public images distortion (Blendon et al. 2006, Miller et al. 2012), payment policies and litigation issues (Charlene et al. 2001, David et al. 2004a,b, Fennell et al. 2010, Smith et al. 2007, Stevenson and David 2003, 2008; William et al. 2014).

In the above healthcare policy research, the ultimate goal is to understand how good services can be delivered to patients. The nursing homes' star rating system is CMS's attempt to implement the quality measures developed in the literature and convey the service quality information to the public in a transparent manner. The number of studies on this rating system is growing since its inception in 2009. Li et al. (2013) studied the nursing home satisfaction rate in Massachusetts and found that incorporating consumer's perspective would improve the CMS nursing home reporting efforts. Konetzka et al. (2015) found that the rating system exacerbates disparities in quality by payer source. To the best of our knowledge, rating inflation issues have not been examined in the literature before.

3.3. Misbehavior Detection

Rating inflation is a typical misbehavior that frequently occurs in system operations. The detailed method used in detecting each type of misbehavior can be different, but the common strategy is to first identify the abnormal phenomenon which cannot be rationalized should the misbehavior not exist, then explore the underlying incentive, usually financial-oriented, driving the phenomenon. Mayzlin et al. (2014) found significant differences in reviews from a given hotel between Expedia and TripAdvisor. Since Expedia only allows its customers to post a review, its posting cost is significantly higher than TripAdvisor, where everyone can post. Consequently, competitors have the incentive to post fake reviews on the "free" TripAdvisor, but not on the "costly" Expedia, and the results gave a good explanation to the observed difference in the two websites' reviews. Duggan and Steven (2000) conducted a study on Japan's elite sumo wrestlers to detect statistical evidence of match rigging, and found that the winning ratio for players on the margin is significantly higher than players who are not. They showed that the incentive structure of promotion leads to gains from trade between wrestlers on the margin for achieving a winning record, and the observed higher winning ratio cannot be simply explained by legitimate effort. Jacob and Steven (2003) studied teachers' cheating behavior using data from Chicago public schools. He found evidence indicating that high-powered incentive systems, especially those with bright line rules, may induce unexpected behavioral distortions such as cheating.

In the above studies, a measure of the abnormal phenomenon, such as review scores, winning ratio, or consistent wrong answer patterns, can be easily accessible. However, due to the illicit nature, the people committing misbehaviors usually attempt not to leave evidence. As a result, sometimes a good measure of the abnormal phenomenon cannot be easily identified, and a good proxy variable is needed to perform the analysis. DellaVigna and La Ferrara (2010) proposed a method to detect illegal arms trade between countries under arms embargo using the weapon manufacturers' stock prices as a proxy and analyzing their fluctuations as turmoil and conflicts arise at certain geographical areas. Engelberg et al. (2014) used the geographic distance between a doctor's office and drug company headquarters to instrument for the likelihood of pecuniary transfers. They found evidence that doctors tilt prescriptions in favor of the paying firm's drugs, shifting away from both branded and generic substitutes.

3.4. Contribution of the Current Research

This research contributes to the existing literature in several dimensions. For operation management literature, nursing homes have not been the research focus despite their importance in the US healthcare system. Different from hospitals and clinics, nursing homes' patients are also residents at the same time, and therefore a lot of efficiency-related performance measures for hospitals and clinics, such as waiting time, readmission rates, do not apply for nursing homes. Many nursing home problems, however, are the results of chronical misbehaviors in the daily care, which may not be objectively measured. Our research results provide a better understanding of the nursing home performance measures and their potential inflation, thus fill up the gap in the nursing home operations management field.

For healthcare policy literature, the existing studies are based on the assumption that the reported data is truth-reflecting and the unbiased results can be delivered to the public. If inflation exists in the rating procedure, then no matter how complete the quality measures are developed or how effective the policies are set, they do not have a truth-reflecting and solid

Please Cite this article in press as: Han, X., et al. Winning at All Costs: Analysis of Inflation in Nursing Homes' Rating System. *Production and Operations Management* (2017), https://doi.org/10.1111/poms.12804

ground, and thus the results will be biased and misleading. Our research targets the authenticity of the reported data and the ratings directly, and provides a solid ground for other research which rely on these data.

Nursing home rating system inflation belongs to the type of problems in which the phenomenon, or the difference between honest nursing homes and inflators, can be difficult to identify, both cross-sectional and longitudinal. The difficulties lie in the following aspects. First, the inflators are confounded with the honest nursing homes whose star ratings also increase after self-reporting, thus ratings cannot be directly used as a measure of inflation. Furthermore, there has been no audit system implemented for the self-reported measures, and there is no data for caught inflators available, which can be used to summarize unique characteristics of inflators. As a result, there is no training data for machine learning-based techniques, making it challenging to design detection methods. From the time dimension perspective, the self-reported measures have been used for years without being audited, thus the rating patterns, although probably inflated, can be very consistent over the years. The lack of external shock also makes it challenging to identify abnormal patterns in the rating data. To overcome these difficulties, we bring in the information from the patients' side, and use the number of complaints as a proxy variable of the true service quality. We then derive contradiction to show that the self-reported rating increase is beyond what can be explained by legitimate efforts. Theoretically, our research provides a framework for detecting rating system inflations: For any product or service to be rated, the ratings are generated from authority inspection, self-reporting, consumer reporting, or their combinations. The three are correlated and can be good proxy variables for justifying others and detecting rating inflation.

Most of the literature only focuses on proving the existence of misbehaviors. However, system reform often takes time, and it is always necessary to give a quantifiable evaluation on the current system's performance. To the best of our knowledge, few studiess have addressed this issue before. Our research contributes to the existing literature by not only demonstrating the existence of rating inflation, but by providing a systematical method to quantify the extent of rating inflation.

4. Financial Incentive Analysis

4.1. Data

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 *Health 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 1219 nursing homes in the state of California over the first 5 years since the inception of the 5-star rating system, that is, from year 2009 to 2013.

The OSHPD data include detailed financial information on California nursing homes over the same period of time. In this dataset, nursing homes' source of revenue is categorized into healthcare and non-healthcare sections. The healthcare section is further classified by revenue source into Medicare, Medicaid,[2] Self-paying, Managed Care, and others. The corresponding revenue and expense details for each section are provided, and the profits can be easily calculated.

The CDPH data are provided through the Health Facilities Consumer Information System (HFCIS) website. A consumer portal is also available on the HFCIS website through which a complaint against a facility can be filed directly. CDPH inspects nursing homes at least once every 6 to 15.9 months in response to these complaints as well as other accidents or incidents that are required to be reported by nursing homes themselves, such as fires, disasters, suspected abuse, etc. Depending on the deficiencies found during the investigation, various types of citations will be issued. A deficiency violating state laws will be issued a state citation, and if it also violates federal law, it will also be reported to CMS and included in the federal inspection for determining star rating. The CDPH data we collected contain detailed patient complaints, which will be used as a proxy of the nursing home's service quality. Note that the state level agency CDPH and the federal level agency CMS, although may overlap sometimes, have independent jurisdictions on nursing home inspections. The CDPH complaints may not be included in CMS's star rating procedure, and the deficiencies covered in CMS's inspection may not result from a CDPH complaint.

4.2. Nursing Homes' Financial Incentive

The observed rating improvement consists of both legitimate efforts and self-reported 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 the 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 combined data have 4433 records for California nursing homes over the 5 years. The average profit per day per patient is calculated for nursing homes in each *Overall* rating group, as shown in Table 3. 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 3 stars in *Health Inspection* may only expect a \$10.79 profit from treating one patient for 1 day. However, if it gains two additional stars after self-reporting and achieves an *Overall* rating of 5 stars, its expected profit increases to \$19.8. Figure 3 shows the profit trend for each of the star rating group over the 5 years.

Nursing homes' total net profits consist of healthcare part and non-healthcare part. The price of healthcare related services is regulated by CMS's Nursing

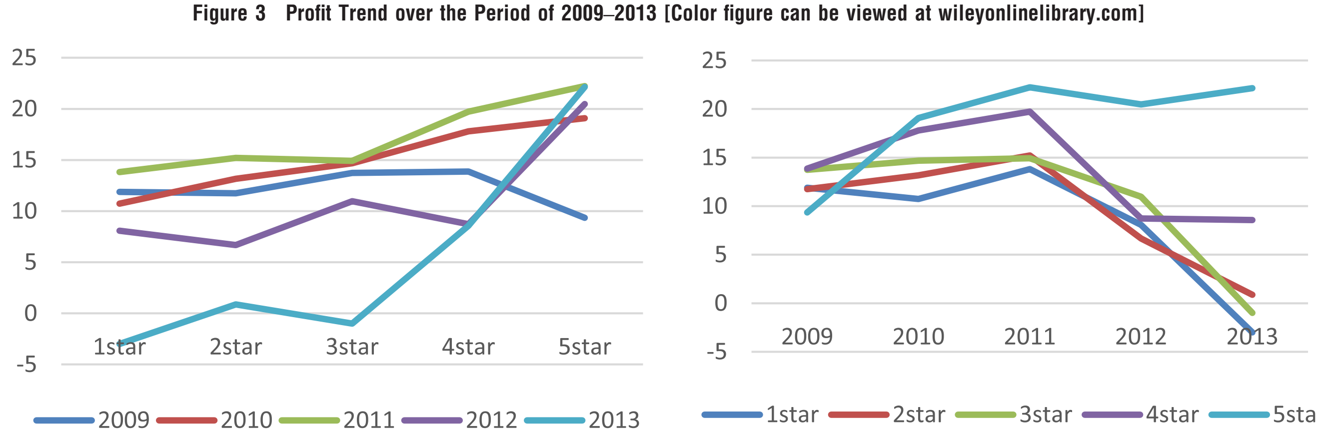
Table 3 Definition of Financial Incentive

Health inspection rating	Expected profit*	Maximum possible Overall rating	Maximum expected profit†	Financial incentive‡
5	19.801	5	19.801	0
4	13.602	5	19.801	6.199 (Level 5 − Level 4)
3	10.790	5	19.801	9.011 (Level 5 − Level 3)
2	10.108	4	13.602	3.494 (Level 4 − Level 2)
1	9.286	2	10.108	0.822 (Level 2 − Level 1)

Notes: *If inspection rating unchanged. The expected profit is the average per patient per day profit for the corresponding star rating group.
†If maximum possible *Overall* rating realized.
‡Difference between expected profit and expected loss.

Home Prospective Payment System (PPS) (CMS, 2017b), which does not consider nursing homes' star ratings. As a result, highly rated nursing homes do not necessarily gain higher healthcare profits than the low-rating nursing homes. The non-healthcare related services, however, are not regulated by CMS. Such services include residential care services, unrestricted contributions, and interest income and gains from investments. Historical data show that highly rated nursing homes can attract more patients who are in good financial conditions (typically self-paying and other resources). These patients are willing to pay more for good quality non-healthcare services. As a result, the non-healthcare profits for high-rating nursing homes can be significantly higher compared with their low-rating counterparts. Moreover, the increased demand for services that happens as a result of high star ratings (Werner et al. 2016) can reduce their overhead costs and thus lead to an increase in the net per-patient profit. 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. 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 3. Note that the financial incentive arises from the expectations in both profits and losses. It is possible for a nursing home that has received 5 stars in *Health 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 perfect *Health Inspection* can be significantly under staffed or provides very poor quality of care. In our dataset, while 125 nursing homes initially rated 3 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.

Figure 3 Profit Trend over the Period of 2009–2013 [Color figure can be viewed at wileyonlinelibrary.com]



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

Please Cite this article in press as: Han, X., et al. Winning at All Costs: Analysis of Inflation in Nursing Homes' Rating System. *Production and Operations Management* (2017), https://doi.org/10.1111/poms.12804

4.3. Empirical Model Specification

We focus on the change in the star rating that happens as a result of the self-reported measures. Our dependent variable, *StarChange*, is equal to the difference between the *Overall* rating and the *Health Inspection* rating. For example, if the nursing home receives 3 stars from *Health Inspection* but receives a 5-star *Overall* rating after including its self-reported measures on *Staffing* and *Quality Measures* 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, α_{-2}, α_{-1}, α_0, α_1, which define the cutoff points of the five levels. *StarChange* for nursing home i at year t can be modeled as follows

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

where $j \in \{-2, -1, 0, 1\}$ and $\boldsymbol{x}$ is a vector of the following independent variables: *Incentive*, *BedCert*, *OccuRate*, *MarketShare*, *HHI*, *ForProfit*, *Medicare*, *Medicaid*, *CouRes*, *CouFam*, *PctgMedicare*, *PctgMedicaid*, *PctgSelfPay*, *PctgMGD*, *Chain*.

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 3 varies depending on the inspection rating of a nursing home. The capacity of each nursing home is measured by the number of certified bed, and is denoted by variable *BedCert*. The occupancy of a nursing home is denoted by variable *OccuRate*, *OccuRate* $\in [0, 1]$. Variables *BedCert* and *OccuRate* together, define the average number of residence of a nursing home. Nursing homes are located in different areas, and may face different market conditions. To capture local market features, we use variable *MarketShare* to denote the market share of each nursing home in its local market, defined by health service area. Based on market share, we also calculate the *Herfindahl–Hirschman Index* (*HHI*), which is widely used for capturing local market competition, and included it in our empirical model. 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 members and get problems resolved more efficiently. Since nursing home residents may be more vulnerable than normal 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. The OSHPD data categorize nursing home payers into five categories: Medicare, Medicaid, Self-Pay, Managed Care, and Others. To capture the impact of different payer percentage on nursing homes' star rating changes, we incorporate the percentage of each type of payers. Four variables, *PctgMedicare*, *PctgMedicaid*, *PctgSelf-Pay*, *PctgMGD* are added to denote the percentage of Medicare payers, Medicaid payers, Self-paying payers, and Managed care payers. The percentage of other type payers are excluded due to multicollinearity. In the nursing home industry, a certain amount of nursing homes is running under some chains. Compared with nursing homes working as separate facilities, nursing homes in chains may have different operational rules and self-reporting behaviors. In our pooled California data, we have over 1500 records of nursing homes in a chain, and there are totally 101 distinct chains. As a result, we do not have sufficient observations for each of the chains to conduct a fixed effect analysis. Rather than adding a chain-level fixed effect, we regroup the nursing homes and add binary variable *chain*, which equals 1 if the nursing home is operating in a chain and 0 if the nursing home is operating separately. Table 4 provides the summary statistics of all variables in our model.

4.4. Estimation Results

We estimate Equation (1) by different methods, as shown in Table 5. The first column shows the estimation results for the pooled data. To deal with potential endogeneity, we take nursing homes' fixed effects into account and run a panel data regression, estimates of which are shown in the second column. Some of the variables in our model are time-invariant. For example, if a nursing home is Medicare certified in the first year, it will most likely remain Medicare certified throughout the following years. As a result, we cannot estimate their coefficients directly through the fixed effect method. To obtain the coefficients of these time-invariant variables, we 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. In all the three models, we observe negative significant coefficients for variable *chain*, indicating that for nursing homes operating in

Table 4 Variable Summary Statistics

Variable	Mean	SD	Minimum	Maximum
Incentive	4.497	3.201	0	9.011
BedCert	101.964	49.579	19	391
OccuRate	0.874	0.172	0.0497	1
ForProfit	0.891	0.292	0	1
Chain	0.767	0.423	0	1
Medicare	0.963	0.188	0	1
Medicaid	0.965	0.183	0	1
CouRes	0.979	0.143	0	1
CouFam	0.230	0.421	0	1
MarketShare	0.0165	0.0159	0.000277	0.125
HHI	5.249	11.271	0.000765	156.314
PctgMedicare	0.154	0.127	0	0.921
PctgMedicaid	0.647	0.235	0	1
PctgSelfPay	0.0838	0.129	0	1
PctgMGD	0.663	0.103	0	0.999

Table 5 Estimates of Equation (1)

Variables	Pooled data	Fixed effect	Hausman–Taylor
Incentive	0.0325***	0.074***	0.074***
	(0.0906)	(0.0144)	(0.0143)
BedCert	0.000284	–	0.00162
	(0.000655)		(0.00207)
OccuRate	−0.584***	0.627	0.445
	(0.166)	(0.560)	(0.468)
ForProfit	−0.0128	–	−1.578***
	(0.112)		(0.332)
Chain	−0.269***	0.0265	−0.444**
	(0.0698)	(0.247)	(0.148)
Medicare	−0.805***	–	−2.684***
	(0.170)		(0.472)
Medicaid	−0.622***	–	−1.323*
	(0.184)		(0.595)
CouRes	−0.231	–	−0.293
	(0.202)		(0.536)
CouFam	−0.279***	–	−0.315
	(0.0693)		(0.174)
MarketShare	−6.79	−71.073*	−60.348*
	(4.49)	(34.363)	(28.568)
HHI	0.020***	0.0777*	0.0781*
	(0.00614)	(0.0343)	(0.0339)
PctgMedicare	−2.283***	4.280***	4.242***
	(0.339)	(1.290)	(1.269)
PctgMedicaid	−0.771***	1.610	1.819
	(0.254)	(1.090)	(1.080)
PctgSelfPay	−0.861***	−4.851***	−4.447***
	(0.321)	(1.063)	(1.042)
PctgMGD	−0.446	6.291***	6.386***
	(0.369)	(1.207)	(1.198)

chains, their star rating increases are less likely to be driven by their financial incentives.

4.5. Alternative Incentive Definition

In the above section, we define the financial incentive of nursing homes based on the average per patient daily profit over the 5-year period. The financial incentive, however, may vary over the years. For example, the difference in the average per-patient daily profit between 3-star nursing homes and 5-star nursing homes in year $t + 1$ may be bigger than that in year t. To capture this change over the years and to test the robustness of our result, we propose an alternative incentive definition in this section. Instead of looking at a 5-year average level, we instead use the per patient daily profit difference of the year t to define nursing homes' financial incentive of year $t + 1$. Table 6 lists the new financial incentive under the new definition. Table 7 then gives the regression results under the alternative financial incentive definition. Similar to the discussion in the previous section, we also run three models: the pooled data model, fixed effect model, and Hausman–Taylor model. In all the three models, the main effect financial incentive is positive significant, which demonstrates the robustness of our results.

5. Inflation Detection and Demonstration

5.1. Correlation Analysis

Although the preliminary results show a positive association between the financial incentive and the changes in the star rating, 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 *Health Inspection* domain 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 *Health Inspection* and those of nursing homes' self-reported domains. Second, if a nursing home really puts an effort in improving its care quality, these efforts should have a lasting effect and lead to better results in the next year's *Health Inspection* and thus there should be a positive correlation between the star ratings from self-reported domains in 1 year and *Health Inspection* ratings in the subsequent year.

Figure 5 shows the two sets of correlations as described above. It can be seen that within the same year, the correlation between *Health Inspection* and *Staffing* is only 0.083, while the correlation between *Health Inspection* and *Quality Measures* is 0.153. The result clearly indicates inconsistency between the *Health Inspection* domain and the self-reported domains within the same year. For the two consecutive years, the correlation between the *Staffing* domain and the *Health Inspection* in the following year is −0.094, and the correlation between the *Quality Measures* domain and the *Health Inspection* in the following year is 0.078. The result indicates that the

Please Cite this article in press as: Han, X., et al. Winning at All Costs: Analysis of Inflation in Nursing Homes' Rating System. *Production and Operations Management* (2017), https://doi.org/10.1111/poms.12804

Table 6 Alternative Definition of Financial Incentive

Health inspection rating	2010 financial incentive	2011 financial incentive	2012 financial incentive	2013 financial incentive
5	0	0	0	0
4	0	0	0.38	3.655
3	0	0.687	3.399	5.183
2	2.129	3.525	3.97	2.858
1	0	1.126	1.447	0.937

Notes: The financial incentive of year t is defined using the year $t-1$ data. Since the panel we collected is from 2009 to 2013, we have no data to define incentives for year 2009, and the year 2013 data (which should be used for 2014 according to the definition) is not used in this definition. In early years (2009 and 2010), there is no significant difference in per patient daily profit for some of the rating levels, thus the financial incentive for improving star rating is defined as 0.

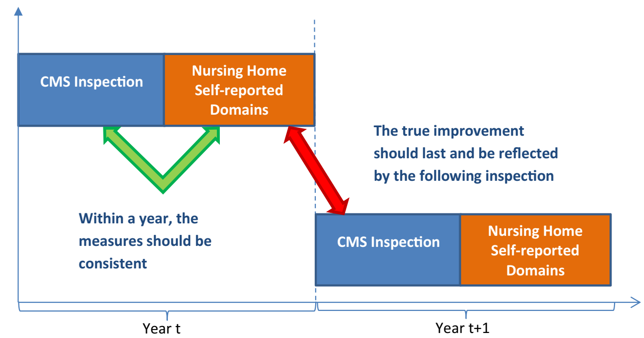
Table 7 Estimates of Equation (1) based on the Alternative Financial Incentive Definition

Variables	Pooled data	Fixed effect	Hausman–Taylor
Incentive	0.323***	0.412***	0.412***
	(0.0202)	(0.0253)	(0.0252)
BedCert	−0.000111	–	−0.000531
	(0.000733)		(0.00210)
OccuRate	−0.515**	0.050	0.257
	(0.190)	(0.625)	(0.517)
ForProfit	−0.0384	–	−0.907**
	(0.126)		(0.352)
Chain	−0.288***	−0.336	−0.522***
	(0.790)	(0.265)	(0.148)
Medicare	−0.900***	–	−2.119***
	(0.189)		(0.486)
Medicaid	−0.461*	–	−1.309*
	(0.203)		(0.623)
CouRes	−0.273	–	−0.244
	(0.227)		(0.495)
CouFam	−0.214**	–	−0.142
	(0.0777)		(0.166)
MarketShare	−8.128	−5.752	−19.306
	(5.101)	(39.159)	(32.179)
HHI	0.0218**	0.0308	0.0344
	(0.00706)	(0.0392)	(0.0388)
PctgMedicare	−2.189***	0.644	0.828
	(0.375)	(1.574)	(1.551)
PctgMedicaid	−0.862**	1.241	1.312
	(0.282)	(1.264)	(1.254)
PctgSelfPay	−0.377	−2.206	−1.990
	(0.364)	(1.281)	(1.248)
PctgMGD	−0.563	3.397*	3.481**
	(0.406)	(1.357)	(1.350)

self-reported improvements in *Quality Measures* and *Staffing* domains have no lasting effect on the next year's *Health Inspection* results at all. The correlation analysis serves as a preliminary evidence of potential inflation, and triggers our further analysis.

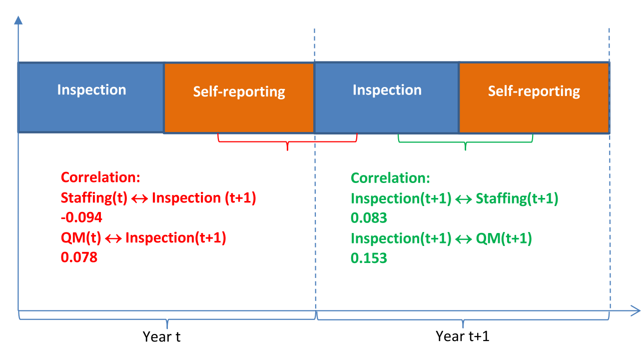
For the 5-year period we analyzed, it appears that the correlation between *Health Inspection* and *Quality Measures* is higher. One interpretation is that the

Figure 4 Graphical Representation of Correlation Analysis [Color figure can be viewed at wileyonlinelibrary.com]



Note: If star increase 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 Correlation Analysis for Five Consecutive Years 2009–2013 [Color figure can be viewed at wileyonlinelibrary.com]



inflation on the *Staffing* domain is relatively easier than that of *Quality Measures* during this period. We do notice that CMS is gradually releasing amendment on the nursing home rating system, and one important policy is to require all nursing homes to report payroll-related staffing data since the beginning of 2016. This shows that the inflation on staffing level is also one of the major concerns of CMS, and once payroll-related staffing data is reported, the *Staffing* measure will become more difficult to be inflated.

5.2. 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 (Anderson et al. 1997, Gardner 2004, Johnson 2001, Rust and Chung 2006). Specifically, we conduct an analysis based on the CDPH complaint data which is an independently collected dataset of patient complaints of California nursing homes. The combined CMS, OSHPD, and CDPH dataset has 3850 records of California nursing homes over the 5 years.

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 similar number of complaints. Table 8a shows the average number of complaints for nursing homes with different *Health Inspection* and *Overall* ratings. In view that larger nursing homes with more patients may get more complaints, we normalized the number of complaints by the size of a nursing home. The normalized results are presented in Table 8b. 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 1. *If the improvements observed did not result from legitimate efforts and inflation does exist, nursing homes with the same Overall star rating but different Health Inspection ratings should have different complaint distributions.*

The results of two ANOVA tests are presented in Table 9a. 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 8 have similar distributions. In the second column, we group nursing homes with the same *Overall* rating based on their *Health Inspection* ratings. In other words, we examine if all the cells in each column of Table 8 have similar distributions. As reported in Table 9a, 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. The ANOVA test results for the normalized complaints are reported in Table 9b, which are similar to the results in Table 9a and support our conclusion.

Table 8 Average Number of Patient Complaints

(a) Original complaints

		Overall star rating				
		1	2	3	4	5
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

(b) Normalized complaints (size = 100)

		Overall star rating				
		1	2	3	4	5
Inspection stars	1	6.505	6.909			
	2	5.946	5.687	6.210	8.251	
	3	4.366	4.676	4.860	5.597	4.288
	4		4.218	4.459	4.157	3.547
	5			9.473	2.517	2.921

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. We denote the shaded and unshaded areas as Area I and Area II, respectively.

Table 9 *F*-statistics: Comparison in Each Overall Rating (a) Original Complaints (b) Normalized Complaints (size = 100)

(a) *F*-statistics

		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***

(b) *F*-statistics

		Grouped by Area I vs. Area II	Grouped by inspection ratings
Overall star rating	1	–	0.94
	2	7.77**	3.25*
	3	7.2**	2.88*
	4	10.15**	5.33**
	5	3.92*	2.64

Claim 2. *If the improvements observed did not result 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 10a. In the first column, nursing homes with the same *Health 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 8 have similar distributions. In the second column, we group

Please Cite this article in press as: Han, X., et al. Winning at All Costs: Analysis of Inflation in Nursing Homes' Rating System. *Production and Operations Management* (2017), https://doi.org/10.1111/poms.12804

Table 10 *F*-statistics: Comparison in Each Inspection Rating (a) Original Complaints (b) Normalized Complaints (size = 100)

(a) *F*-statistics

		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

(b) *F*-statistics

		Grouped by Area I vs. Area II	Grouped by Overall ratings
Inspection ratings	1	0.5	0.5
	2	1.6	1.46
	3	0.99	0.77
	4	2.51	0.95
	5	–	3.3*

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 8 have similar distributions. As shown in Table 10a, 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 2 is also supported. Together with the results obtained for Claim 1, the analysis provides strong evidence of the existence of rating inflation in self-reported measures. The ANOVA test results for the normalized complaints are reported in Table 10b, which also support our conclusion.

6. 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 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.

6.1. 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 II, *OverallRating* is ordinal and takes values in five levels $\{1, 2, \ldots, 5\}$, 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

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

where $k \in \{1, 2, 3, 4\}$. The independent variables denoted by vector $\boldsymbol{x}$ 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.

Our logic for selecting the honest nursing homes is as follows: We divide the inflators into two different types. The first type consists of nursing homes which inflate their self-reported measures to achieve higher ratings. For these inflators to be identified, a necessary condition is that they gain additional stars after self-reporting (there can be honest nursing homes who gain the additional stars through legitimate efforts though). In our first criterion, we excluded all the nursing homes whose star rating increased after self-reporting, thus we completely excluded any inflators of this type. The second type of inflators consists of nursing homes which inflate their self-reported measures to avoid losing stars. These nursing homes may have low *Staffing* or low *Quality Measures* that may lead to decreased *Overall* ratings. In our second criterion, we excluded nursing homes whose number of complaints are above the median of its rating level. By doing this, we excluded nursing homes that may lose stars due to their poor services, and guaranteed that the remaining nursing homes deserve staying in that rating level.

Based on the two criteria, we identify the honest (*H*) group, which consists of 1262 nursing home records in 5 years. The remaining 2588 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.

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. The regression results for the honest group is reported in Table 11. Both the 95% and 90% confidence interval are calculated. 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% confidence interval, we can identify 147 inflator records out of the 2588 nursing home records (5.68%) in the PI group. Based on 90% confidence interval, we can identify 219 inflator records (8.46%) in the PI group.

We observe that nursing homes with 4-star Health Inspection ratings compose the largest proportion (41.6%) of the likely inflators identified. A possible explanation is that it is easier for these 4-star nursing homes to cover up their "slightly" inflated self-reporting, and it is less likely to trigger a red flag in case of an audit.

Table 11 Estimates of the Honest Group

Variables	Coefficients
Incentive	0.102***
	(0.0184)
BedCert	−0.0115***
	(0.00125)
OccuRate	0.863**
	(0.300)
ForProfit	−1.491***
	(0.208)
Chain	−0.430***
	(0.130)
Medicare	−2.853***
	(0.563)
Medicaid	−1.580***
	(0.408)
CouRes	−0.613
	(0.420)
CouFam	0.646***
	(0.122)
MarketShare	2.234
	(8.146)
HHI	0.00889
	(0.0116)
PctgMedicare	−0.330
	(0.758)
PctgMedicaid	−0.0678
	(0.626)
PctgSelfPay	1.582*
	(0.718)
PctgMGD	−2.551***
	(0.793)

6.2. 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 likely inflators nor identified as honest ones. The remaining dataset consists of 1481 nursing home records, in which 1262 records are for honest nursing homes and 219 records are for the likely inflators identified using a 90% confidence interval. The status of a nursing home is assigned as 0 if it belongs to honest group and 1 if it is a likely inflator. To perform the variable importance analysis, we use the logistic specification presented in Equation (3).

$$\mathrm{logit}(\lambda) = \boldsymbol{x}'\boldsymbol{\beta} \tag{3}$$

where λ is the probability of being identified as an inflator, $\boldsymbol{x}$ is the vector of variables that were also used in Equations (1) and (2).

The variable importance analysis results are presented in Table 12. Among the variables, we find the variable *BedCert* 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 percentage of self-paying is also a key variable contributing to being an inflator. As discussed earlier in the incentive definition section, self-paying patients are typically good in

Please Cite this article in press as: Han, X., et al. Winning at All Costs: Analysis of Inflation in Nursing Homes' Rating System. *Production and Operations Management* (2017), https://doi.org/10.1111/poms.12804

Table 12 Variable Importance Analysis

Variables	Variable importance
BedCert	0.405
Pctg_SelfPay	0.288
Cou_Fam	0.274
Chain	0.264
Incentive	0.254
Occu_Rate	0.194
ForProfit	0.182
HHI	0.176
MarketShare	0.171
Pctg_MGD	0.122
Medicaid	0.109
Pctg_Medicaid	0.097
Cou_Res	0.066
Pctg_Medicare	0.043
Medicaid	0.041

financial situations and they contribute significantly to nursing homes' non-healthcare profits. Since non-healthcare pricing is not regulated by CMS, highly rated nursing homes typically charge much higher prices on non-healthcare services than low-rating nursing homes. It is reasonable to believe that many nursing homes with high percentage of self-paying patients are inflating their self-reported measures in order to gain more non-healthcare profits. The results also indicate that nursing homes with family type councils are more likely to be inflators. Another variable *Chain* also has a very high importance. Note that in section 4, our results suggest that nursing homes in chains are less likely to be driven by their financial incentives to improve star ratings. One possible explanation for these results is that nursing homes in these franchises follow chain-level decisions for self-reporting, which are less sensitive to individual nursing home's financial incentive. It is possible for some chains to inflate self-reported measures throughout their facilities. Besides the variables discussed above, *Incentive* and *ForProfit* are also important factors for being an inflator. For-profits 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 result is consistent with the work of Chesteen et al. (2005). The probability of being an inflator, on the other hand, is less likely to be affected by market competition (e.g., *HHI*, *Market Share*) and certification status (*Medicare*, *Pctg_Medicare*, *Medicaid*, *Pctg_Medicaid*, etc.).

7. Conclusion

This study 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 with stronger financial incentives are more likely to improve their star ratings after self-reporting, either through improving service quality or through inflating self-reported measures. We then develop a systematical method which uses 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 key characteristics of 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 the observed behavior. Second, we contribute to the theory by developing a systematical method for demonstrating the existence of rating inflation and evaluating the inflator proportion. As we discussed earlier, although CMS has announced amendment policies to improve the rating system, the rating generation mechanism still heavily depends on the self-reported measures, and thus the issue of inflation still exists. This research demonstrates the shortcoming of the current rating system and provides a guideline for CMS on how to improve the rating system. The characteristics summarized for the likely inflators provide valuable information for CMS to develop an effective audit system in the future.

This work also has several limitations. First, we are unable to measure the financial incentive for each nursing home at individual level. This is practically very difficult, since even for the same nursing home staying in the same rating level, the financial incentive may vary over the time due to various financial situations. To address this limitation, we perform our analysis on an aggregated level for each rating group, and use the group average as the incentive for nursing homes in the same group, leaving the unobserved incentive fluctuations to the nursing homes fixed effects. Second, since CMS does not have an audit system in place, there is no data for detected inflators available for us to evaluate the performance of our prediction model. Such difficulties caused by data limitation are a common issue in misbehavior detection research. We address this limitation by calculating the highest possible rating using confidence intervals and giving conservative predictions. Third, we are only able to measure patient complaints in numbers, but not in "severeness." For example, a complaint on medical malpractice may have much more impact than a complaint on sanity. Future research may consider applying text mining techniques to address this limitation.

Acknowledgment

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

Notes

[1]Additional conditions apply to nursing homes which are in the CMS's Special Focus Facility program.
[2]In California, Medicaid is referred to as Medi-Cal. However, we use Medicaid as the category name in this study, in order to avoid confusion for readers from other states.

References

Anderson, E., F. Claes, R. Roland. 1997. Customer satisfaction, productivity, and profitability: Differences between goods and services. *Market. Sci.* **16**(2): 129–145.

Anderson, D., G. Gao, B. Golden. 2014. Life is all about timing: An examination of differences in treatment quality for trauma patients based on hospital arrival time. *Prod. Oper. Manag.* **23** (12): 2178–2190.

Andritsos, D. A., S. Aflaki. 2015. Competition and the operational performance of hospitals: The role of hospital objectives. *Prod. Oper. Manag.* **24**(11): 1812–1832.

Bauer, T., S. Mathias. 2008. An extension of the Blinder–Oaxaca decomposition to nonlinear models. *AStA Adv. Stat. Anal.* **92** (2): 197–206.

Bercovitz, A., L. L. Dwyer, A. Jones, G. W. Strahan. 2009. The national nursing home Survey: 2004 overview. *Vital Health Stat. Ser.* **13**: 1–155.

Berg, K., V. Mor, J. Morris, K. Murphy, T. Moore, Y. Harris. 2002. Identification and evaluation of existing nursing homes quality indicators. *Healthcare Financ. Rev.* **23**(4): 19–36.

Blendon, R. J., M. Brodie, J. M. Benson, D. E. Altman, T. Buhr. 2006. Americans' views of healthcare costs, access, and quality. *Milbank Q.* **84**(4): 623–657.

Carman, J. M. 1990. Consumer perceptions of service quality: An assessment of the SERVQUAL dimensions. *J. Retail.* **66**(1): 33–55.

Castle, N. G. 2008. Nursing home caregiver staffing levels and quality of care a literature review. *J. Appl. Gerontol.* **27**(4): 375–405.

Cayirli, T., E. Veral. 2003. Outpatient scheduling in health care: A review of literature. *Prod. Oper. Manag.* **12**(4): 519–549.

Centers for Medicare and Medicaid Services. 2015. Nursing home compare 3.0: Revisions to the nursing home compare 5-star quality rating system. Centers for Medicare and Medicaid Services. https://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2015-Fact-sheets-items/2015-02-12-2.html (accessed date September 22, 2017).

Centers for Medicare and Medicaid Services. 2017a. Design for nursing home compare five-star quality rating system: Technical users' guide. Report, Centers for Medicare and Medicaid Services, Baltimore, MD.

Center for Medicare and Medicaid Services. 2017b. Skilled nursing facility PPS. Centers for Medicare and Medicaid Services. https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/SNFPPS/index.html (accessed date September 22, 2017).

Charlene, H., S. Woolhandler, J. Mullan, H. Carrillo, D. Himmelstein. 2001. Does investor ownership of nursing homes compromise the quality of care? *Am. J. Public Health* **91**(9): 1452–1455.

Chen, H., Q. Qian, A. Zhang. 2015. Would allowing privately funded health care reduce public waiting time? Theory and empirical evidence from canadian joint replacement surgery data. *Prod. Oper. Manag.* **24**(4): 605–618.

Chesteen, S., B. Helgheim, T. Randall, D. Wardell. 2005. Comparing quality of care in non-profit and for-profit nursing homes: A process perspective. *J. Oper. Manag.* **23**(2): 229–242.

Chow, V. S., M. L. Puterman, N. Salehirad, W. Huang, D. Atkins. 2011. Reducing surgical ward congestion through improved surgical scheduling and uncapacitated simulation. *Prod. Oper. Manag.* **20**(3): 418–430.

Dabholkar, P. A., D. I. Thorpe, J. O. Rentz. 1995. A measure of service quality for retail stores: Scale development and validation. *J. Acad. Mark. Sci.* **24**(1): 3–16.

David, G., J. Angelelli, V. Mor. 2004a. Medicaid payment and risk-adjusted nursing home quality measures. *Health Aff.* **23**(5): 243–252.

David, G., Z. Feng, O. Intrator, V. Mor. 2004b. Recent trends in state nursing home payment policies. *Health Aff.* **W4**: 363–373.

DellaVigna, S., E. La Ferrara. 2010. Detecting illegal arms trade. *Am. Econ. J.* **2**(4): 26–57.

Dobson, G., S. Hasija, E. J. Pinker. 2011. Reserving capacity for urgent patients in primary care. *Prod. Oper. Manag.* **20**(3): 456–473.

Duggan, M., L. Steven. 2000. Winning isn't everything: Corruption in sumo wrestling. National Bureau of Economic Research Working Paper No. 7798.

Duhigg, C. 2007. Washington scrutinizes nursing homes. *The New York Times*. Available at http://www.nytimes.com/2007/11/16/business/16care.html (accessed date September 22, 2017).

Eiken, S., K. Sredl, L. Gold, J. Kasten, B. Burwell, P. Saucier. 2014. Medicaid expenditures for long-term services and supports in FFY 2012. Report, Truven Health Analytics, Cambridge, MA.

Engelberg, J., P. Christopher, T. Nathan. 2014. Financial conflicts of interest in medicine. Social Science Research Network. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2297094 (accessed date September 22, 2017).

Fairlie, R. 2005. An extension of the Blinder–Oaxaca decomposition technique to logit and probit models. *J. Econ. Social Measure.* **30**(4): 305–316.

Fennell, M., Z. Feng, M. Clark, V. Mor. 2010. Elderly hispanics more likely to reside in poor-quality nursing homes. *Health Aff.* **29**(1): 65–73.

Fowles, D. G. 2012. A profile of older Americans: 2012. Report, Administration on Aging, U.S. Department of Health and Human Services, Washington, DC.

Gardner, E. 2004. Dimensional analysis of airline quality. *Interfaces* **34**(4): 272–279.

General Accounting Office. 2002. Nursing homes: Public reporting of quality indicators has merit, but national implementation is premature. Report, U.S. General Accounting Office, Washington, DC.

Gokula, R. R. M., J. A. Hickner, M. A. Smith. 2004. Inappropriate use of urinary catheters in elderly patients at a midwestern community teaching hospital. *Am. J. Infect. Control* **32**(4): 196–199.

Grabowski, D. C., A. J. O'Malley. 2014. Use of telemedicine can reduce hospitalizations of nursing home residents and generate savings for medicare. *Health Aff.* **33**(2): 244–250.

Güneş, E. D., E. L. Örmeci, D. Kunduzcu. 2015. Preventing and diagnosing colorectal cancer with a limited colonoscopy resource. *Prod. Oper. Manag.* **24**(1): 1–20.

Harrington, C., D. Zimmerman, S. L. Karon, J. Robinson, P. Beutel. 2000. Nursing home staffing and its relationship to deficiencies. *J. Gerontol Ser. B Psychol. Sci. Soc. Sci.* **55**(5): S278–S287.

Harrington, C., B. Olney, H. Carrillo, T. Kang. 2012. Nurse staffing and deficiencies in the largest for-profit nursing home chains and chains owned by private equity companies. *Health Serv. Res.* **47**(1pt1): 106–128.

Helm, J. E., S. AhmadBeygi, M. P. Van Oyen. 2011. Design and analysis of hospital admission control for operational effectiveness. *Prod. Oper. Manag.* **20**(3): 359–374.

Jack, E. P., T. L. Powers. 2004. Volume flexible strategies in health services: A research framework. *Prod. Oper. Manag.* **13**(3): 230–244.

Jacob, B., L. Steven. 2003. Rotten apples: An investigation of the prevalence and predictors of teacher cheating. National Bureau of Economic Research Working Papers No. 9413.

Johnson, R. 2001. Linking complaint management to profit. *Int. J. Serv. Ind. Manag.* **12**(1): 60–69.

Kane, R. L., R. A. Kane. 2001. What older people want from long-term care, and how they can get it. *Health Aff.* **20**(6): 114–127.

Kim, H., C. Kovner, C. Harrington, W. Greene, M. Mezey. 2009. A panel data analysis of the relationships of nursing home staffing levels and standards to regulatory deficiencies. *J. Gerontol. Ser. B* **64**(2): 269–278.

Konetzka, R. T., D. Yi, E. C. Norton, K. E. Kilpatrick. 2004. Effects of Medicare payment changes on nursing home staffing and deficiencies. *Health Serv. Res.* **39**(3): 463–488.

Konetzka, R. T., D. C. Grabowski, M. C. Perraillon, R. M. Werner. 2015. Nursing home 5-star rating system exacerbates disparities in quality, by payer source. *Health Aff.* **34**(5): 819–827.

LaGanga, L. R., S. R. Lawrence. 2012. Appointment overbooking in health care clinics to improve patient service and clinic performance. *Prod. Oper. Manag.* **21**(5): 874–888.

Li, Y., X. Cai, Z. Ye, L. G. Glance, C. Harrington, D. B. Mukamel. 2013. Satisfaction with massachusetts nursing home care was generally high during 2005–09, with some variability across facilities. *Health Aff.* **32**(8): 1416–1425.

Mayzlin, D., D. Yaniv, C. Judith. 2014. Promotional reviews: An empirical investigation of online review manipulation. *Am. Econ. Rev.* **104**(8): 2421–2455.

McCoy, J. H., M. Eric Johnson. 2014. Clinic capacity management: Planning treatment programs that incorporate adherence. *Prod. Oper. Manag.* **23**(1): 1–18.

Miller, E. A., D. A. Tyler, J. Rozanova, V. Mor. 2012. National newspaper portrayal of US nursing homes: Periodic treatment of topic and tone. *Milbank Q.* **90**(4): 725–761.

Mor, V., K. Berg, J. Angelelli, D. Gifford, J. Morris, T. Moore. 2003. The quality of quality measurement in US nursing homes. *Gerontologist* **43**(suppl 2): 37–46.

Mor, V., J. Zinn, J. Angelelli, J. M. Teno, S. C. Miller. 2004. Driven to tiers: Socioeconomic and racial disparities in the quality of nursing home care. *Milbank Q.* **82**(2): 227–256.

Mor, V., O. Intrator, Z. Feng, D. C. Grabowski. 2010. The revolving door of rehospitalization from skilled nursing facilities. *Health Aff.* **29**(1): 57–64.

Munroe, D. J. 1990. The influence of registered nurse staffing on the quality of nursing home care. *Res. Nurs. Health* **13**(4): 263–270.

Oaxaca, R. 1973. Male–female wage differentials in urban labor markets. *Int. Econ. Rev.* **14**(3): 693–709.

Rosalie, K. 2003. Definition, measurement, and correlates of quality of life in nursing homes: Toward a reasonable practice, research, and policy agenda. *Gerontologist* **43**(suppl 2): 28–36.

Roth, A. V., R. Van Dierdonck. 1995. Hospital resource planning: Concepts, feasibility, and framework. *Prod. Oper. Manag.* **4**(1): 2–29.

Rust, R., T. T. S. Chung. 2006. Marketing models of service and relationships. *Market. Sci.* **25**(6): 560–580.

Salzarulo, P. A., S. Mahar, S. Modi. 2015. Beyond patient classification: Using individual patient characteristics in appointment scheduling. *Prod. Oper. Manag.* **25**(6): 1056–1072.

Shugarman, L. R., J. A. Brown. 2006. Nursing home selection: How do consumers choose? Volume I: Findings from focus groups of consumers and information intermediaries. Report, Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health & Human Services, Washington, DC.

Shwartz, M., J. D. Restuccia, A. K. Rosen. 2015. Composite measures of health care provider performance: A description of approaches. *Milbank Q.* **93**(4): 788–825.

Smith, D. B., Z. Feng, M. Fennell, J. Zinn, V. Mor. 2007. Separate and unequal: Racial segregation and disparities in quality across US nursing homes. *Health Aff.* **26**(5): 1448–1458.

Stavropoulou, C., C. Doherty, P. Tosey. 2015. How effective are incident-reporting systems for improving patient safety? A systematic literature review *Milbank Q.* **93**(4): 826–866.

Stevenson, D. 2006. Is a public reporting approach appropriate for nursing home care? *J. Health Polit. Policy Law* **31**(4): 773–810.

Stevenson, D., S. David. 2003. The rise of nursing home litigation: Findings from a national survey of attorneys. *Health Aff.* **22**(2): 219–229.

Stevenson, D., G. David. 2008. Private equity investment and nursing home care: Is it a big deal? *Health Aff.* **27**(5): 1399–1408.

Thomas, K. 2014. Medicare star ratings allow nursing homes to game the system. *The New York Times*. Available at https://www.nytimes.com/2014/08/25/business/medicare-star-ratings-allow-nursing-homes-to-game-the-system.html (accessed date September 22, 2017).

Tsaur, S. H., T. Y. Chang, C. H. Yen. 2002. The evaluation of airline service quality by Fuzzy MCDM. *Tourism Manage.* **23**(2): 107–115.

Turnham, H. 2001. Federal nursing home reform act from the omnibus budget reconciliation act of 1987 or simply OBRA '87. National Long-Term Care Ombudsman Resource Center, Washington, DC.

Walshe, K. 2001. Regulating US nursing homes: Are we learning from experience? *Health Aff.* **20**(6): 128–144.

Werner, R., T. Konetzka. 2010. Advancing nursing home quality through quality improvement itself. *Health Aff.* **29**(1): 81–86.

Werner, R., E. Stuart, D. Polsky. 2010. Public reporting drove quality gains at nursing homes. *Health Aff.* **29**(9): 1706–1713.

Werner, R. M., R. Tamara Konetzka, D. Polsky. 2016. Changes in consumer demand following public reporting of summary quality ratings: An evaluation in nursing homes. *Health Serv. Res.* **51**(S2): 1291–1309.

William, C., D. Himmelstein, M. Siman, S. Woolhandler. 2014. For-profit medicare home health agencies' costs appear higher and quality appears lower compared to nonprofit agencies. *Health Aff.* **33**(8): 1460–1465.

Zepeda, E. D., K. K. Sinha. 2016. Toward an effective design of behavioral health care delivery: An empirical analysis of care for depression. *Prod. Oper. Manag.* **25**(5): 952–967.

Zhang, X., D. C. Grabowski. 2004. Nursing home staffing and quality under the nursing home reform act. *Gerontologist* **44**(1): 13–23.

Zhang, Y., M. L. Puterman, M. Nelson, D. Atkins. 2012. A simulation optimization approach to long-term care capacity planning. *Oper. Res.* **60**(2): 249–261.

Please Cite this article in press as: Han, X., et al. Winning at All Costs: Analysis of Inflation in Nursing Homes' Rating System. *Production and Operations Management* (2017), https://doi.org/10.1111/poms.12804