

# Impact of Public Reporting on Unreported Quality of Care

*Rachel M. Werner, R. Tamara Konetzka, and Gregory B. Kruse*

---

**Objective.** The impact of quality improvement incentives on nontargeted care is unknown and some have expressed concern that such incentives may be harmful to nontargeted areas of care. Our objective is to examine the effect of publicly reporting quality information on unreported quality of care.

**Data Sources/Study Setting.** The nursing home Minimum Data Set from 1999 to 2005 on all postacute care admissions.

**Study Design.** We studied 13,683 skilled nursing facilities and examined how unreported aspects of clinical care changed in response to changes in reported care after public reporting was initiated by the Centers for Medicare and Medicaid Services on their website, Nursing Home Compare, in 2002.

**Principal Findings.** We find that overall both unreported and reported care improved following the launch of public reporting. Improvements in unreported care were particularly large among facilities with high scores or that significantly improved on reported measures, whereas low-scoring facilities experienced no change or worsening of their unreported quality of care.

**Conclusions.** Public reporting in the setting of postacute care had mixed effects on areas without public reporting, improving in high-ranking facilities, but worsening in low-ranking facilities. While the benefits of public reporting may extend beyond areas that are being directly measured, these initiatives may also widen the gap between high- and low-quality facilities.

**Key Words.** Quality of care, postacute care, nursing home quality, public reporting

---

Public reporting of quality information is a potentially powerful tool to improve health care quality. Moving quality information into the public domain may improve quality of care by giving consumers the information necessary to choose high-quality providers. Additionally, health care providers may respond to this information by improving the quality of care they provide. Because improving quality in this way is theoretically appealing and the potential for a positive effect on quality is substantial, public reporting is increasingly being adopted for hospitals, health plans, nursing homes, home health agencies, and physicians.

While public reporting has the potential to improve quality of care in areas that are being measured and reported, one potential limitation is the impossibility of measuring *all* the important aspects of care. By necessity, measures of clinical care are limited to what is measurable, and what is measurable is not always what is most important. The impact of quality improvement incentives on unreported care is unknown, and some policy makers and clinicians have expressed concern that there may be unintended and negative consequences to public reporting, including causing providers to focus their attention on measured aspects of care while neglecting unmeasured but important areas of care (Casalino 1999; Werner and Asch 2005).

Our objective is to examine the effect of publicly reporting quality information on unreported quality of care. We do this in the setting of Nursing Home Compare, a public reporting initiative launched by the Centers for Medicare and Medicaid Services (CMS) in 2002 to address quality deficits in nursing homes. Using clinical measures of quality, Nursing Home Compare publicly rates all Medicare- and Medicaid-certified nursing homes in the United States on the care they provide for short-stay and chronic-care residents. In this setting, we examine how unreported aspects of clinical care changed in response to public reporting of other aspects of care.

## PRIOR EVIDENCE

Little prior work has examined the impact of performance measurement on unmeasured quality of care. One randomized, controlled trial (Mohide et al. 1988) examined the impact of a quality improvement intervention in nursing homes on areas of care that were not targeted by the intervention. The study found that while care for the targeted conditions improved, there was no change in the care for the nontargeted condition. More recently, Asch et al. (2004) examined the effects of performance measurement in the Veterans Health Administration (VHA) on targeted and nontargeted conditions. For

---

Address correspondence to Rachel M. Werner, M.D., Ph.D., 1230 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104; e-mail: [rwerner@mail.med.upenn.edu](mailto:rwerner@mail.med.upenn.edu). Dr. Werner is with the Center for Health Equity Research and Promotion, Philadelphia VAMC, Division of General Internal Medicine, University of Pennsylvania School of Medicine, Leonard Davis Institute of Health Economics, University of Pennsylvania. R. Tamara Konetzka, Ph.D., is with the Department of Health Studies, University of Chicago, Chicago, IL. Gregory B. Kruse, M.Sc., M.P.H., is with the Department of Health Care Systems, The Wharton School, University of Pennsylvania, Locust Walk, Philadelphia, PA.

areas of care that were targeted by performance measurement within the VHA, they found that patients from the VHA were more likely to receive recommended care than non-VHA patients. However, the difference in care between VHA and non-VHA patients in conditions that were not targeted by VHA performance measurement were smaller and barely reached statistical significance. Another recent study tested whether a quality improvement intervention changed nontargeted care for vulnerable elders in an ambulatory care setting (Ganz et al. 2007). This observational study of a practice redesign intervention found that while targeted care processes improved in the intervention practices compared with control practices, there was no change in the nontargeted care processes in either practice setting. Finally, one prior study has tested the effect of quality improvement incentives on targeted and nontargeted care, examining a hospital pay-for-performance program for acute myocardial infarction (Glickman et al. 2007). The study found that neither targeted nor nontargeted care processes significantly changed in response to pay-for-performance.

Our study contributes to this existing literature in two important ways. First, only one prior study has examined changes in nontargeted quality of care in the face of market-based quality improvement incentives (Glickman et al. 2007), but in finding no improvement in targeted measures, it provides little evidence of how nontargeted care changes when health care providers improve care in targeted areas. While other work has not found a significant change in nontargeted care, quality improvement from market-based incentives such as public reporting of quality may provide stronger incentives for improving targeted quality and therefore stronger potential for effects on nontargeted quality. Second, prior work has not directly correlated changes in targeted care to changes in nontargeted care *within* health care providers. While others have answered the question, “Does nontargeted quality change *on average?*” in this study we ask the question, “Does nontargeted quality change *in response to changes in targeted quality?*”

## METHODS

### *Conceptual Framework*

Holmstrom and Milgrom’s (1991) theory of multitasking predicts that measuring and rewarding quality in some areas may harm quality in other areas. This is specifically the case when quality is multidimensional and when quality improvement efforts target only some dimensions of quality.

In the setting of health care, market-based quality improvement incentives, such as public reporting and pay for performance, typically reward only a subset of all measures. In addition, some aspects of health care quality are difficult to measure. Thus, large segments of health care quality are currently unrewarded, and in some cases unmeasured. Because quality is multidimensional, multitasking theory predicts that providers will divert resources away from these unrewarded and unmeasured aspects of quality.

The degree to which rewarded and unrewarded measures evaluate the same dimension of care may predict whether unrewarded measures improve in response to improvements in rewarded measures. If rewarded and unrewarded quality measures are related to the same quality dimension, we may expect that efforts focused on improving quality tied to incentives will spill over to unrewarded areas, causing both to improve. Conversely, measures related to different quality dimensions may be more likely to diverge when incentives are related to only one measure, as focusing limited resources on rewarded care may crowd out unrewarded care.

The approach to quality improvement may also predict whether unrewarded care improve in response to improvements in rewarded areas. If improvements in quality are driven by structural changes, such as by hiring more professional nursing staff, we may expect both rewarded and unrewarded areas of care to improve (to the extent that both are related to nurse staffing). Alternatively, if improvements in quality are driven by targeted changes, such as changes in protocols and work organization, resources may be diverted away from unrewarded quality, resulting in worsening of unrewarded quality while rewarded quality improves.

### *Empirical Approach*

We test the effect of public reporting on unrewarded care in the setting of public reporting of nursing home quality, or Nursing Home Compare. In 2002, the CMS launched the Nursing Home Quality Initiative, an effort to improve quality of care in nursing homes. Working with measurement experts, the National Quality Forum, and a diverse group of nursing home industry stakeholders, CMS adopted a set of nursing home quality measures. CMS publicly released this information in Nursing Home Compare, a web-based resource (<http://www.medicare.gov/NHCompare>) that details quality of care at all Medicare- or Medicaid-certified nursing homes. Nursing Home Compare began rating nursing homes based on 10 quality measures, three of which evaluate the quality of postacute care. On November 12, 2002, the publication of these quality

measures was launched nationally, allowing consumers to compare nursing home quality measures across 17,000 nursing homes nationwide (Centers for Medicare and Medicaid 2002; Harris and Clauser 2002).

We focus our analysis on the quality measures for skilled nursing facilities (SNF) related to postacute care, because a substantial number of measures of postacute care quality have been validated but not reported in Nursing Home Compare, and test whether reported and unreported aspects of care change after the launch of Nursing Home Compare in November 2002. To test whether changes in unreported care are related to quality improvement on reported measures, we stratify our analyses by whether SNFs improved or were high scoring on reported measures after Nursing Home Compare was launched. Finally, we test for changes in nurse staffing after the launch of Nursing Home Compare, as we expect increases in nurse staffing to have positive spillovers onto unrewarded aspects of care.

### *Data*

The primary data source for our analyses is the nursing home Minimum Data Set (MDS) for years 1999–2005. The MDS contains detailed clinical data that is collected at regular intervals for every resident in a Medicare- or Medicaid-certified nursing home. Data on residents' health, physical functioning, mental status, and psycho-social well-being have been collected electronically since 1998. These data are used by nursing homes to assess the needs of and develop a plan of care unique to each resident and by the CMS to calculate Medicare prospective reimbursement rates. A recent large field reliability trial of MDS was conducted to verify the quality of the data. Research nurses shown to be highly reliable among themselves undertook up to 30 reliability assessments in each of 209 randomly selected facilities. Results reveal that over 85 percent of MDS data elements have adequate interrater reliability ( $\kappa > 0.6$ ) and those below that threshold were very low prevalence binary indicators showing high levels of agreement (Mor et al. 2003). Other researchers have found considerable evidence for the consistency of clinical data in the instrument (Gambassi et al. 1998). Because of the reliability of these data and the detailed resident-level clinical information contained therein, they are considered the best available data for measuring nursing home quality and thus are the source for the quality measures reported on Nursing Home Compare.

### *Quality Measures*

Calculating the quality measures directly from the MDS allows us to produce the quality measures both before and after Nursing Home Compare was

released. We use the technical definitions of the quality measures provided by CMS (Morris et al. 2003) to calculate each nursing home's quality measures for postacute care residents over the time period of the study. We calculated all the postacute care quality measures that were publicly reported on Nursing Home Compare when it was launched in November 2002: percent of short-stay residents who did not have moderate or severe pain, percent of short-stay residents without delirium, and percent of short-stay residents whose walking improved. We followed all the conventions of the CMS quality measures to calculate the postacute care measures: each measure is calculated quarterly over two-quarters of data on all residents with a 14-day MDS assessment; only those residents who stay in the facility long enough to have a 14-day assessment are included in the calculation of the aggregate quality measure; only facilities with at least 20 cases during the target time period are included. After calculating the measures, our results were benchmarked against the publicly reported quality measures that are available from CMS. The results of our calculations for these quality measures differed from the results reported by CMS between 0 and 0.15 percentage points, indicating an excellent match between the results of our measurement calculations and those published by CMS.

We also used MDS to calculate postacute care quality measures identified by Abt Associates Inc. as valid, but not chosen by CMS to be publicly reported. In 2004, Abt assessed the validity of a large number of measures of the quality of postacute care (Moore et al. 2005), classifying quality measures as Level I (highest validity), Level II (moderate validity), and Level III (not valid). The technical details of this validation process have been previously described (Moore et al. 2005). For the purposes of our study, we included all the unreported quality measures with Level I or Level II validity. We calculated nine valid measures of quality of care that were not reported in Nursing Home Compare—eight with Level I validity and one with Level II validity<sup>1</sup>—based on the technical definitions for these measures (Moore et al. 2005).

While we expect two of the measures of unreported quality (improved pain and locomotion) that are closely related to two reported quality dimensions (pain and walking) to change with reported quality, current evidence does not lead us to make predictions about the extent to which the remaining unreported measures may be related to or share production resources with reported measures.

All the quality measures were rescaled so that a higher score indicates higher quality of care. Reported and unreported quality measures are summarized in Table 1.

Table 1: Summary Statistics for Reported and Unreported Quality Measures for Postacute Care Residents

<i>Measures</i>	<i>Description</i>	<i>Mean (SD)</i>
Reported		
Pain	% of residents who did not have moderate or severe pain	75.8 (14.6)
Delirium	% of residents without delirium	96.2 (5.4)
Walking	% of residents whose walking improved	7.7 (7.2)
Unreported		
Improved pain	% of residents whose pain improve or remained free from pain	52.8 (14.8)
Locomotion	% of residents whose level of locomotion functioning remained independent or improved	32.4 (14.5)
Shortness of breath	% of residents who do not have shortness of breath	82.5 (11.7)
Bladder incontinence	% of residents who improved their bladder incontinence or remained fully continent	49.2 (14.3)
Respiratory infection	% of residents who did not develop a respiratory infection or had a respiratory infection that got better	94.9 (5.1)
UTI	% of residents without a urinary tract infection (UTI)	77.1 (10.4)
ADL	% of residents with improving level of activities of daily living (ADL) functioning	50.6 (19.8)
Mid-loss ADL	% of residents who improve status on mid-loss ADL functioning (transfer or locomotion) or remain independent in mid-loss ADLs	40.2 (16.0)
Early loss ADL	% of residents who improve status on early loss ADL functioning (dressing and personal hygiene) or remain completely independent on early loss ADLs	27.9 (15.4)

*Analyses*

To test for changes in reported and unreported quality when Nursing Home Compare was launched, we used facility-level linear regressions where each quality measure (reported and unreported) was a function of Nursing Home Compare indicator variables, time-varying covariates, and SNF fixed effects. We tested for changes in quality measures with the launch of Nursing Home Compare in November 2002 in two ways: (1) using year indicator variables (2001–2005, omitting 2000) we tested whether quality changed at the launch of Nursing Home Compare, between 2002 and 2003; and (2) using a pre–post indicator variable (2000–2002 versus 2003–2005), we tested whether the

quality level differed in the pre-Nursing Home Compare period compared with the post-Nursing Home Compare period. We included the following time-varying covariates to control for the significant changes the case mix of patients being admitted to postacute care during our study time period: the mean Cognitive Performance Scale (Morris et al. 1994), the mean activities of daily living (ADL) summary score (a measure of ADL) (Morris et al. 1999), and the percentage of SNF residents admitted in each Resource Utilization Group (RUG). Because all the quality measures are calculated only on residents who remain in postacute care for at least 14 days, we also control for the “censoring” rate at each facility using quarterly measures of the proportion of all postacute care admissions that remain in postacute care for 14 days at each SNF. In all the cases, robust standard errors were used to account for nonindependence of observations from the same facility.

To specifically test how unreported quality changed in response to changes in reported quality of care, we stratified the above analyses of unreported quality by whether facilities improved or were high ranking on reported measures. We defined facilities that improved or were high ranking on reported measures as those that improved on all the three reported measures between 2001 and 2005 plus facilities that scored in the top one-third on all the three reported measures after Nursing Home Compare was launched. Conversely, facilities that did not improve or were low ranking on reported measures were defined as those that stayed the same or worsened on all the three reported measures between 2001 and 2005 plus facilities that scored in the lower one-third on all the three reported measures after Nursing Home Compare was launched.

We also examined changes in nurse staffing over the study period to investigate whether structural improvements may be responsible for positive spillovers to unreported quality of care, calculating nurse staffing measures from On-Line Survey Certification and Reporting System. We followed standard procedures to calculate staffing ratios, assuming that each full-time equivalent staff member works 70 hours in a 2-week period and dividing the staffing hours per day by the number of residents in the facility (Abt Associates Inc. 2001). We measured skilled staffing intensity as registered nurse (RN) plus licensed practical nurse (LPN) hours per resident day.

We checked the robustness of our results to model choice. In the base model described above, we used ordinary least squares. However, because our dependent variable is a proportion, and thus bounded between 0 and 1, it may violate the assumptions of OLS. This is particularly the case for quality measures that contain a large number of zeros or ones. We test the sensitivity



of our results to this possibility in two ways. First, we transform our dependent variable using the logit transformation, which results in a normally distributed dependent variable, and re-estimate these regressions using OLS. While this approach allows us to adequately test for the statistical significance of unbiased estimates, it introduces the need to retransform the dependent variable, making interpretation of the results less straightforward. Therefore, we also use generalized estimating equations with binomial family and logit link. This approach is attractive because the link function directly characterizes how the expectation on the original scale is related linearly to the predictors, which overcomes problems with retransformation of the dependent variable (Papke and Wooldridge 1996). However, this approach produces a matrix-weighted average of the between-SNF and within-SNF effects, rather than the within-SNF effects that are the goal of this analysis.

We also re-estimated our base model on a balanced panel of SNFs present through all the 6 years of the study ( $n = 8,225$  facilities). By excluding facilities that either dropped out of the panel or entered the panel in a later time period, we can confirm that the effect we observe of Nursing Home Compare on reported and unreported quality is due to changes that occurred within the same group of facilities rather than changes in the composition of the panel.

## RESULTS

Multivariate within-SNF analyses showed that all the three reported measures of quality improved after the launch of Nursing Home Compare (Table 2). Between 2002 and 2003, when Nursing Home Compare was launched, the percentage of patients whose pain was controlled improved by 2.6 percentage points (on a base of 76 percent), the percentage of patients without delirium improved by 0.5 percentage points (on a base of 96 percent), and the percentage of patients with improved walking improved by 0.4 percentage points (on a base of 8 percent). Improvements in the quality measures averaged over the postperiod were larger.

Several of the unreported measures also improved in 2003 immediately after Nursing Home Compare was launched (Table 2): improved pain (2.5 percentage points on a base of 53 percent), locomotion (0.3 percentage points on a base of 32 percent), shortness of breath (0.6 percentage points on a base of 83 percent), and bladder incontinence (0.6 percentage points on a base of 49 percent). These improvements persisted throughout the postperiod.

**Table 2: Changes in Reported and Unreported Measures after Nursing Home Compare (NHC) Was Launched in 2002**

	<i>Reported Measures</i>		
	<i>Pain</i>	<i>Delirium</i>	<i>Walking</i>
2000 (omitted)			
2001	-0.00591*** (0.00096)	0.00928*** (0.00051)	0.000756 (0.00050)
2002	0.00250*** (0.0012)	0.0150*** (0.00058)	0.00282*** (0.00056)
2003	0.0281*** (0.0014)	0.0199*** (0.00062)	0.00658*** (0.00061)
2004	0.0295*** (0.0014)	0.0224*** (0.00063)	0.0105*** (0.00068)
2005	0.0290*** (0.0015)	0.0254*** (0.00064)	0.0146*** (0.00073)
Constant	0.708*** (0.0083)	0.980*** (0.0037)	0.283*** (0.0051)
Observations	235262	232484	216581
Number of facilities	13674	13601	13198
$R^2$	0.03	0.06	0.09
Change at implementation of NHC (between 2002 and 2003)	0.0256***	0.00486***	0.00377***
Change between pre-NHC (2000–2002) and post-NHC (2003–2005)	0.0294***	0.0139***	0.00863***

		<i>Unreported Measures</i>								
		<i>Improved Pain</i>	<i>Locomotion</i>	<i>Shortness of Breath</i>	<i>Bladder Incontinence</i>	<i>Respiratory Infection</i>	<i>UTI</i>	<i>ADL</i>	<i>Mid-Loss ADL</i>	<i>Early-Loss ADL</i>
2000 (omitted)										
2001		-0.0113*** (0.0011)	0.00403*** (0.0011)	0.00628*** (0.00085)	0.00520*** (0.00093)	0.00389*** (0.00046)	0.0000339 (0.00080)	-0.00279** (0.0014)	0.00262** (0.0012)	-0.00435*** (0.0012)
2002		-0.00820*** (0.0012)	0.00548*** (0.0013)	0.00528*** (0.0010)	0.00614*** (0.0011)	0.00548*** (0.00049)	-0.00483*** (0.00092)	-0.0108*** (0.0017)	-0.000322 (0.0014)	-0.0123*** (0.0014)
2003		0.0169*** (0.0014)	0.00889*** (0.0014)	0.0112*** (0.0011)	0.0123*** (0.0012)	0.00225*** (0.00052)	-0.00739*** (0.00097)	-0.0203*** (0.0018)	0.000578 (0.0015)	-0.0206*** (0.0015)
2004		0.0147*** (0.0015)	0.00852*** (0.0015)	0.0137*** (0.0012)	0.0171*** (0.0012)	0.00352*** (0.00054)	-0.00912*** (0.0010)	-0.0325*** (0.0019)	-0.00974*** (0.0016)	-0.0299*** (0.0016)
2005		0.00429*** (0.0015)	0.00198 (0.0016)	0.0174*** (0.0012)	0.0174*** (0.0013)	0.00603*** (0.00054)	-0.0120*** (0.0011)	-0.0474*** (0.0020)	-0.0219*** (0.0017)	-0.0435*** (0.0016)
Constant		0.638*** (0.0099)	0.300*** (0.010)	0.857*** (0.0074)	0.912*** (0.010)	0.956*** (0.0037)	0.902*** (0.0065)	0.436*** (0.013)	0.320*** (0.011)	0.309*** (0.010)
Observations		22118	218266	235214	233487	221250	235274	215229	218254	218263
Number of facilities		1334	13254	13675	13620	1336	13675	13137	13254	13254
R <sup>2</sup>		0.03	0.00	0.02	0.19	0.01	0.03	0.02	0.01	0.03
Change at implementation of NHC (between 2002 and 2003)		0.0251*** (0.0015)	0.00341*** (0.0016)	0.00592*** (0.0012)	0.00619*** (0.0013)	-0.00323*** (0.00054)	-0.00255*** (0.0011)	-0.000946*** (0.0020)	0.000900 (0.0017)	-0.000835*** (0.0016)
Change between pre-NHC (2000-2002) and post-NHC (2003-2005)		0.0189*** (0.0015)	0.00368*** (0.0016)	0.0105*** (0.0012)	0.0111*** (0.0013)	0.000918*** (0.00054)	-0.00902*** (0.0011)	-0.0268*** (0.0020)	-0.00973*** (0.0017)	-0.0240*** (0.0016)

\* $p < .1$ ; \*\* $p < .05$ ; \*\*\* $p < .01$ ;

Robust standard errors in parentheses. Covariates: cognitive performance scale; ADL, activities of daily living summary scale; NHC, Nursing Home Compare; RUG, Resource Utilization Groups; UTI, urinary tract infection.

The measure of respiratory infections worsened slightly immediately after Nursing Home Compare in 2003 (−0.3 percentage points on a base of 95 percent), but then improved again and was on average better in the 3 years after Nursing Home Compare was launched compared with before. Several unreported measures worsened in 2003 immediately after Nursing Home Compare was launched: urinary tract infection (−0.2 percentage points on a base of 77 percent), ADL functioning (−0.9 percentage points on a base of 51 percent), and early loss ADLs (−0.8 percentage points on a base of 28 percent). Quality in these areas steadily trended downward from 2000 through 2005, suggesting these declines may not be associated with Nursing Home Compare.

To test whether changes in unreported measures were related to changes in reported measures, we stratified the analyses based on whether or not a SNF was high scoring on reported measures (Table 3). In general, we found facilities that were high scoring on reported measures improved on unreported measures. Specifically, among high-scoring facilities, the unreported measures that were related to the same quality dimension as reported measures (improved pain and locomotion) had gains in quality, as did several other measures (shortness of breath and bladder incontinence). At the same time, facilities that were low scoring on reported measures had significantly smaller improvements, had no significant change, or worsened on unreported measures after Nursing Home Compare was launched. For unreported measures that worsened on average, the decrement in quality was generally larger among low-scoring facilities than high-scoring facilities.

Finally, we examined changes in nurse staffing related to Nursing Home Compare, as nurse staffing in one way nursing homes may improve quality of care that may lead to an improvement in unreported quality, rather than a worsening. We focus on professional nurse staffing (RNs and LPNs) because skilled services are by definition of fundamental importance for postacute care in SNF. We find that professional nurse hours per resident-day declined over the study period (Table 4), for reasons that may have to do with financial pressures (Konetzka et al. 2004) and that we assume are unrelated to public reporting. However, the relative declines in professional staffing after Nursing Home Compare are consistently smaller for high-scoring facilities than for low-scoring facilities. The maintenance of relatively higher staffing in high-scoring facilities is consistent with the generally larger positive spillovers to unreported quality found at these facilities.

Our robustness checks confirmed that our choice of model—linear regression with facility fixed effects—was not driving our results. Our findings

Table 3: Changes in Unreported Measures, Stratified by Facilities That Are High-Scoring versus Low-Scoring on Reported Measures

	Improved Pain		Locomotion		Shortness of Breath		Bladder Incontinence	
	High-Scoring Facilities	Low-Scoring Facilities	High-Scoring Facilities	Low-Scoring Facilities	High-Scoring Facilities	Low-Scoring Facilities	High-Scoring Facilities	Low-Scoring Facilities
2000 (omitted)								
2001	0.000624 (0.0024)	-0.0285*** (0.0034)	0.0106*** (0.0025)	-0.00166 (0.0036)	0.0134*** (0.0020)	-0.00104 (0.0028)	0.00700*** (0.0020)	0.00452 (0.0029)
2002	0.0163*** (0.0027)	-0.0428*** (0.0040)	0.0136*** (0.0029)	0.000423 (0.0044)	0.0191*** (0.0023)	-0.0137*** (0.0035)	0.00667*** (0.0024)	0.00563 (0.0036)
2003	0.0533*** (0.0032)	-0.0346*** (0.0046)	0.0178*** (0.0031)	-0.00145 (0.0049)	0.0267*** (0.0025)	-0.0132*** (0.0038)	0.0152*** (0.0025)	0.00665* (0.0037)
2004	0.0561*** (0.0032)	-0.0326*** (0.0047)	0.0221*** (0.0033)	-0.00699 (0.0053)	0.0308*** (0.0026)	-0.0103*** (0.0042)	0.0161*** (0.0027)	0.0140*** (0.0039)
2005	0.0485*** (0.0032)	-0.0492*** (0.0048)	0.0148*** (0.0035)	-0.0123** (0.0055)	0.0374*** (0.0027)	-0.00695 (0.0042)	0.0120*** (0.0029)	0.0113*** (0.0041)
Constant	0.610*** (0.021)	0.603*** (0.034)	0.310*** (0.024)	0.218*** (0.034)	0.853*** (0.017)	0.888*** (0.026)	0.856*** (0.022)	0.757*** (0.026)
Observations	41119	19650	40825	19467	42846	20398	42642	20325
Number of facilities	1943	951	1943	951	1943	951	1943	951
R <sup>2</sup>	0.08	0.04	0.01	0.01	0.04	0.02	0.26	0.20
Change at implementation of NHC (between 2002 and 2003)	0.037***	0.00818**	0.00415*	-0.00187	0.00755***	0.000492	0.00853***	0.00103
Change between pre-NHC (2000-2002) and post-NHC (2003-2005)	0.047***	-0.0149***	0.0103***	-0.00512	0.0211***	-0.00482*	0.00931***	0.00619***

	Respiratory Infection		UTI		ADL		Mid-Loss ADL		Early-Loss ADL	
	High-Scoring Facilities	Low-Scoring Facilities	High-Scoring Facilities	Low-Scoring Facilities	High-Scoring Facilities	Low-Scoring Facilities	High-Scoring Facilities	Low-Scoring Facilities	High-Scoring Facilities	Low-Scoring Facilities
2000 (omitted)										
2001	0.00278*** (0.0010)	0.00383*** (0.0016)	0.000754 (0.0018)	-0.00148 (0.0027)	-0.00648*** (0.0031)	0.00156 (0.0046)	0.00266 (0.0027)	0.000940 (0.0039)	-0.00472** (0.0026)	-0.00540 (0.0039)
2002	0.00464*** (0.0011)	0.00661*** (0.0016)	-0.00436*** (0.0021)	-0.00825*** (0.0032)	-0.0165*** (0.0037)	-0.0127*** (0.0054)	0.000233 (0.0032)	-0.00427 (0.0047)	-0.0104*** (0.0031)	-0.0183*** (0.0045)
2003	0.000843 (0.0012)	0.00287* (0.0017)	-0.00380* (0.0022)	-0.0184*** (0.0034)	-0.0265*** (0.0040)	-0.0197*** (0.0060)	-0.00374 (0.0034)	-0.00715 (0.0052)	-0.0177*** (0.0032)	-0.0273*** (0.0050)
2004	0.00369*** (0.0012)	0.00328* (0.0018)	-0.00601*** (0.0023)	-0.0169*** (0.0035)	-0.0396*** (0.0043)	-0.0333*** (0.0065)	-0.00488 (0.0037)	-0.0213*** (0.0057)	-0.0288*** (0.0035)	-0.0364*** (0.0055)
2005	0.00530*** (0.0012)	0.00479*** (0.0018)	-0.00403* (0.0024)	-0.0216*** (0.0036)	-0.0566*** (0.0044)	-0.0463*** (0.0070)	-0.0179*** (0.0038)	-0.0293*** (0.0060)	-0.0427*** (0.0036)	-0.0450*** (0.0038)
Constant	0.959*** (0.0089)	0.963*** (0.013)	0.923*** (0.015)	0.902*** (0.025)	0.403*** (0.031)	0.476*** (0.047)	0.320*** (0.026)	0.289*** (0.039)	0.303*** (0.024)	0.308*** (0.038)
Observations	41131	19663	42851	20401	40314	19337	40824	19467	40827	19467
Number of facilities	1943	951	1943	951	1941	951	1943	951	1943	951
R <sup>2</sup>	0.01	0.01	0.03	0.04	0.03	0.02	0.01	0.01	0.03	0.03
Change at implementation of NHC (between 2002 and 2003)	-0.00379***	-0.00374***	0.000564	-0.0101***	-0.0100***	-0.00701*	0.00351	-0.00288	-0.00730***	-0.00900**
Change between pre-NHC (2000- 2002) and post-NHC (2003-2005)	0.00107	0.000697	-0.00445***	-0.0173***	-0.0319***	-0.0278***	-0.00656***	-0.0163***	-0.0233***	-0.0277***

\*  $p < .1$ ; \*\*  $p < .05$ ; \*\*\*  $p < .01$ ;

Robust standard errors in parentheses.

Covariates: cognitive performance scale; ADL, activities of daily living summary scale; NHC, Nursing Home Compare. RUG, resource utilization groups; UTI, urinary tract infection. High-scoring facilities: facilities that improved on all the three reported measures between 2001 and 2005 or scored in the top one-third on all the three reported measures after Nursing Home Compare was released.

Low-scoring facilities: facilities that stayed the same or worsened on all the three reported measures between 2001 and 2005 or scored in the bottom one-third on all the three reported measures after Nursing Home Compare was released.

Table 4: Changes in Professional Nurse Staffing before versus after Nursing Home Compare (NHC) Was Released

	<i>RNs+ LPNs</i>		
	<i>All Facilities</i>	<i>High-Scoring Facilities</i>	<i>Low-Scoring Facilities</i>
2000 (omitted)			
2001	- 0.00120 (0.0041)	- 0.0157* (0.0082)	- 0.00605 (0.014)
2002	- 0.0214*** (0.0070)	- 0.0332*** (0.015)	- 0.0492** (0.023)
2003	- 0.0164** (0.0071)	- 0.0461*** (0.015)	- 0.0563** (0.024)
2004	- 0.0208*** (0.0073)	- 0.0542*** (0.014)	- 0.0725*** (0.027)
2005	- 0.0148** (0.0071)	- 0.0382*** (0.015)	- 0.0374 (0.027)
Constant	1.284*** (0.037)	1.240*** (0.085)	1.485*** (0.16)
Observations	225035	41551	19678
Number of facilities	13316	1936	947
$R^2$	0.00	0.00	0.00
Change at implementation of NHC (between 2002 and 2003)	0.00500	- 0.0128	- 0.00706
Change between pre-NHC (2000–2002) and post-NHC (2003–2005)	- 0.0107**	- 0.0304***	- 0.0388**

Robust standard errors in parentheses.

\*\*\* $p < 0.01$ .

\*\* $p < 0.05$ .

\* $p < 0.1$ .

Covariates: cognitive performance scale; ADL, activities of daily living summary scale; RUG, resource utilization groups.

did not change qualitatively using a linear regression with logit transformation of the dependent variable, generalized estimating equations, or a balanced panel of SNFs.

## DISCUSSION

There has been significant concern that quality improvement incentives such as public reporting may inadvertently harm care that is not directly targeted by

the incentive. In the setting of public reporting of postacute care, we find that overall both unreported and reported care improved following the launch of public reporting. Improvements in unreported care were particularly large among facilities with high scores or that significantly improved on reported measures, whereas low-scoring facilities experienced no change or worsening of their unreported quality of care.

Our findings among high-ranked facilities are more consistent with the theory that the positive effect of public reporting spills over into other important but unreported areas of nursing home care and less consistent with the theory that measuring and reporting quality in some areas crowds out quality in other areas. The positive effect of public reporting on unreported quality suggests that public reporting may have induced SNFs to make structural and organizational changes that had a positive effect on unreported quality.<sup>2</sup> Our staffing results are consistent with this conclusion. Within the context of a secular decline in professional nurse staffing ratios over the study period, high-scoring facilities were able to maintain higher professional nurse staffing ratios than low-scoring facilities, facilities that in some cases had worsening in unreported quality of care. These findings lend support to the role of nurse staffing in changes in both reported and unreported quality of care.

On the other hand, we found that facilities that were low scoring on reported measures also failed to improve on unreported measures and, in some cases, worsened on these measures. This could be due to a lack of resources, knowledge, or will to pursue any type of quality improvement initiative, or a failed effort to improve on reported measures that nonetheless drew resources from unreported quality. Thus, the main unintended consequence of public reporting of quality in postacute care may not be a growing divide between reported and unreported aspects of quality but rather a growing divide between providers that are more and less able to achieve quality improvement.

The relative changes in unreported quality are, in most cases, smaller than the changes observed in reported quality. Most of these changes are small in magnitude (the absolute improvements in most unreported quality measures are one percentage point or less, which translates into a relative improvement of 1–2 percent). There are multiple reasons for small improvements in quality, including measurement error and loss to follow up, biasing the results toward the null. By contrast, the relative improvements in reported quality are larger. The smaller changes we find in reported quality compared with unreported quality suggests that there were both spillovers to unreported



quality from structural quality improvement efforts and that some resources were directed specifically toward improving reported quality.

Our study has several limitations. As with any observational study, there are potential sources of bias that could impact our results. It is difficult to definitively attribute changes in quality (both reported and unreported) to Nursing Home Compare using a pre-post design. Because we study the national launch of Nursing Home Compare, it is impossible to know whether observed changes would have occurred in the absence of Nursing Home Compare. There may also be unmeasured SNF characteristics that are associated with changes in both reported and unreported care other than Nursing Home Compare. For example, good management may lead to better scores on both reported and unreported measures. To account for this, we include SNF-level fixed effects so that we estimate within SNF changes in quality, accounting for time-invariant differences in SNF. In addition, we include a set of SNF-level control variables to account for time-varying changes in SNF related to changes in patient case mix; however, bias from omitted time-varying attributes may remain. It is possible that the relationship we demonstrate between reported and unreported quality is due to changes in the accuracy of the data rather than true changes in quality of care. While other work has found that changes of this nature explain some quality improvement (Green and Wintfeld 1995; Roski et al. 2003), this is less likely to happen in the nursing home quality measures calculated from the MDS. Electronic MDS data collection started in 1998, long before Nursing Home Compare was launched, and has been used to determine Medicare payment since that time, increasing nursing homes' incentive to accurately report these data for several years before the launch of Nursing Home Compare. Finally, we do not measure changes in overall quality of care, but rather individual elements of unreported care. Thus, our analysis is subject to the global limitation of measuring quality inherent in Nursing Home Compare and similar interventions—individual measures may not represent overall quality, reported or unreported.

Our findings provide good news for quality improvement incentives such as public reporting and pay-for-performance in that fears of serious “crowding out” of unreported quality do not appear to be substantiated in many SNFs. Reported and unreported quality within SNFs generally improved together after public reporting was initiated, particularly in high-ranking SNFs. These findings suggest that investments in quality improvement in the face of market-based incentives, if successful, may improve other areas of care as well. At the same time, lack of improvement on both reported and unreported measures was also common. Fears that

public reporting will lead to a growing divide between high- and low-quality providers deserves further investigation.

## ACKNOWLEDGMENTS

*Joint Acknowledgments/Disclosure Statement:* This research was funded by a grant from the Agency for Healthcare Research and Quality (R01 HS016478-01) and the University Research Foundation of the University of Pennsylvania. Rachel Werner is funded in part by a VA HSR&D Career Development Award. This project is funded, in part, under a grant from the Pennsylvania Department of Health. The Department specifically disclaims responsibility for any analyses, interpretations, or conclusions.

*Disclosures:* None.

*Disclaimers:* None.

*Prior dissemination:* This research was presented at AcademyHealth, Washington, DC, June 2008 and American Society of Health Economics, Durham NC, June 2008.

## NOTES

1. All the unreported measures included in this study had Level I validity except for “percent of short-stay residents with urinary tract infections,” which had Level II validity.
2. It may also be that simply knowing that facility quality would be under public scrutiny provided an impetus for facilities to direct resources to general quality improvement efforts.

## REFERENCES

- Abt Associates Inc. 2001. *Report to Congress: Appropriateness of Minimum Nurse Staffing Ratios in Nursing Homes Phase II Final Report*. Baltimore: Centers for Medicare and Medicaid Services.
- Asch, S. M., E. A. McGlynn, M. M. Hogan, R. A. Hayward, P. Shekelle, L. Rubenstein, J. Keeseey, J. Adams, and E. A. Kerr. 2004. “Comparison of Quality of Care for Patients in the Veterans Health Administration and Patients in a National Sample.” *Annals of Internal Medicine* 141: 938–45.
- Casalino, L. P. 1999. “The Unintended Consequences of Measuring Quality on the Quality of Medical Care.” *New England Journal of Medicine* 341: 1147–50.

- Centers for Medicare and Medicaid. 2002, 1/20/2004. "Nursing Home Quality Initiatives Overview" [accessed on January 14, 2006]. Available at: <http://www.cms.hhs.gov/NursingHomeQualityInits/downloads/NHQIOverview.pdf>
- Gambassi, G., F. Landi, L. Peng, C. Brostrup-Jensen, K. Calore, J. Hiris, L. Lipsitz, V. Mor, and R. Bernabei. 1998. "Validity of Diagnostic and Drug Data in Standardized Nursing Home Resident Assessments: Potential for Geriatric Pharmacoepidemiology. SAGE Study Group. Systematic Assessment of Geriatric Drug Use via Epidemiology." *Medical Care* 36: 167-79.
- Ganz, D. A., N. S. Wenger, C. P. Roth, C. J. Kamberg, J. T. Chang, C. H. MacLean, R. T. Young, D. H. Solomon, T. Higashi, L. Min, D. B. Reuben, and P. G. Shekelle. 2007. "The Effect of a Quality Improvement Initiative on the Quality of Other Aspects of Health Care: The Law of Unintended Consequences?" *Medical Care* 45: 8-18.
- Glickman, S. W., F.-S. Ou, E. R. DeLong, M. T. Roe, B. L. Lytle, J. Mulgund, J. S. Rumsfeld, W. B. Gibler, E. M. Ohman, K. A. Schulman, and E. D. Peterson. 2007. "Pay for Performance, Quality of Care, and Outcomes in Acute Myocardial Infarction." *Journal of the American Medical Association* 297: 2373-80.
- Green, J., and N. Wintfeld. 1995. "Report Cards on Cardiac Surgeons. Assessing New York State's Approach." *New England Journal of Medicine* 332: 1229-32.
- Harris, Y., and S. B. Clauser. 2002. "Achieving Improvement through Nursing Home Quality Measurement." *Health Care Financing Review* 23: 5-18.
- Holmstrom, B., and P. Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization* 7: 24-52.
- Konetzka, R. T., D. Yi, E. C. Norton, and K. E. Kilpatrick. 2004. "Effects of Medicare Payment Changes on Nursing Home Staffing and Deficiencies." *Health Services Research* 39: 463-88.
- Mohide, E. A., P. X. Tugwell, P. A. Caulfield, L. W. Chambers, C. W. Dunnett, S. Baptiste, J. R. Bayne, C. Patterson, K. V. Rudnick, and M. Pill. 1988. "A Randomized Trial of Quality Assurance in Nursing Homes." *Medical Care* 26: 554-65.
- Moore, T., N. Wu, D. Kidder, B. Bell, D. Warner, L. Hadden, Y. Mackiernan, J. Morris, and R. Jones. 2005. "Design and Validation of Post-Acute Care Quality Measures." Abt Associates Subcontract: RAND Prime Contract No. 500-00-0029 (TO 2), Abt Associates Inc.
- Mor, V., J. Angelelli, R. Jones, J. Roy, T. Moore, and J. Morris. 2003. "Inter-Rater Reliability of Nursing Home Quality Indicators in the U.S." *BMC Health Services Research* 3: 20.
- Morris, J. N., B. E. Fries, D. R. Mehr, C. Hawes, C. Phillips, V. Mor, and L. A. Lipsitz. 1994. "MDS Cognitive Performance Scale." *Journal of Gerontology* 49: M174-82.
- Morris, J. N., B. E. Fries, and S. A. Morris. 1999. "Scaling ADLs within the MDS." *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 54: M546-53.
- Morris, J. N., T. Moore, R. Jones, V. Mor, J. Angelelli, K. Berg, C. Hale, S. Morriss, K. M. Murphy, and M. Rennison. 2003. *Validation of Long-Term and Post-Acute*

*Care Quality Indicators*. Baltimore: Centers for Medicare and Medicaid Services.

- Papke, L. E., and J. M. Wooldridge. 1996. "Econometric Methods for Fractional Response Variables with an Application to 411(k) Plan Participation Rates." *Journal of Applied Econometrics* 11: 619–32.
- Roski, J., R. Jeddloh, L. An, H. Lando, P. Hannan, C. Hall, and S. H. Zhu. 2003. "The Impact of Financial Incentives and a Patient Registry on Preventive Care Quality: Increasing Provider Adherence to Evidence-Based Smoking Cessation Practice Guidelines." *Preventive Medicine* 36: 291–9.
- Werner, R. M., and D. A. Asch. 2005. "The Unintended Consequences of Publicly Reporting Quality Information." *Journal of the American Medical Association* 293: 1239–44.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.