Performance of Multihospital Health Systems' Flagship Hospitals in the CMS Star Rating Program

Jianhui Hu, PhD*, David R Nerenz, PhD

Center for Health Policy & Health Services Research, Henry Ford Health System, Detroit, Michigan.

Using the Hospital Compare overall hospital quality star ratings and other publicly available data on acute care hospitals, we examined star ratings for the flagship hospitals of a set of multihospital health systems in the United States. We compared star ratings and hospital characteristics of flagship and nonflagship hospitals across and within 113 health systems. The system flagship hospitals had significantly lower star ratings than did nonflagship hospitals, and they did not generally have the highest star ratings in their own systems. Higher teaching intensity, larger bed size, higher uncompensated care, and higher disproportionate share hospital (DSH) patient percentage were all significantly associated with lower star ratings of flagship hospitals when compared with nonflagship hospitals across all health systems; the flagship hospital of a system was more likely to have the lowest star rating in its system if the difference in DSH percentage was relatively large between the flagship and nonflagship hospitals in that system. *Journal of Hospital Medicine* 2020;15:407-410. © 2020 Society of Hospital Medicine

The Centers for Medicare & Medicaid Services (CMS) Hospital Compare overall hospital ratings was originally released in 2016 and was recently updated in February 2019.[1,2] The program is designed to provide a consumer-friendly global rating system for hospitals, with hospitals rated on a scale from one star (worst) to five stars (best). The ratings are based on a formula that combines scores on 57 performance measures into seven groups, with the groups of mortality, safety, readmission, and patient experience given weights of 22% each in the overall scoring, and groups of effectiveness of care, timeliness of care, and efficient use of medical imaging equally contributing to the rest of the score.

Concerns have been raised since the introduction of the program regarding the methodology and possible unfairly high or low star ratings for certain types of hospitals.[3,4] It has been noted that five-star hospitals are disproportionately small, specialty-focused hospitals that may not have Emergency Departments or significant volumes of Medicaid patients.[5] Hospitals that report fewer measures and thus receive scores for fewer measure groups (in general, smaller or specialty hospitals) are more likely to receive higher star ratings than are hospitals that receive scores for all measure groups.[6,7] Teaching hospitals, on average, have received lower star ratings than nonteaching hospitals.[8,9]

Multihospital systems generally designate one of their hospitals as a "flagship" hospital and often use the name of that hospital to identify the system as a whole (eg, Mayo Clinic Health System, University of Pittsburgh Medical Center). There is not a set of objective criteria to designate a "flagship" hospital of a multihospital health system. Flagships could be the founding hospitals of the systems or the largest hospitals in the systems, and they are usually (although not always) large teaching hospitals. There is therefore a potential paradox in which a set of hospitals that tend to get lower ratings in the CMS star rating system may also be the set frequently identified as system flagship hospitals and whose reputation is used as a brand identity for multihospital systems.

It is possible, though, that the hospitals designated as flagship hospitals in multihospital systems are exceptions to the general rule of lower star ratings for major teaching hospitals. The flagship designation may reflect excellence that is then reflected in the star rating system, or it may reflect some other kind of excellence (eg, reputation for research or teaching, diverse medical services provided) that is not reflected in the star rating system. The primary aim of this study was to compare the average star ratings and hospital characteristics of designated flagship hospitals in multihospital systems with those of (1) major teaching hospitals generally and (2) "nonflagship" hospitals across and within the same systems specifically. We sought to determine whether a flagship designation would be associated with higher star ratings than those of major teaching hospitals in general and with higher star ratings than other, nonflagship hospitals in the same system.

The use of a prestigious flagship hospital name to identify a multihospital system suggests that some aspects of high quality in the flagship are extended in some way to other hospitals in the system. If that is so, then the star ratings of hospitals in organized multihospital systems with a flagship may be more similar to each other than those of sets of hospitals selected at

*Corresponding Author: Jianhui Hu, PhD; Email: jhu1@hfhs.org; Telephone: 313-874-5454; Twitter: @jianhui_hu.

random. As a secondary aim, to determine whether this type of consistent quality throughout a system could be identified in the CMS hospital star rating system, we compared the variation in star ratings between organized multihospital systems with flagship hospitals to those of artificially created "pseudo systems" of unaffiliated hospitals.

METHODS

We used the Agency for Healthcare Research and Quality (AHRQ) Compendium of U.S. Health Systems, 2016, database and hospital file to identify multihospital health systems and their member hospitals.[10] The database also provides information about health system characteristics such as systemwide teaching intensity and total number of acute care hospitals. We linked the AHRQ files to the CMS Hospital Compare datasets and Hospital Inpatient Prospective Payment System (IPPS) 2018 Final Rule Impact File to obtain star ratings and other information about specific hospitals (eg, resident to bed ratio, uncompensated care payment). Throughout the study, we followed the AHRQ's definition of "major teaching hospitals" as hospitals with a high resident to bed ratio (≥0.25).

For purposes of this study, the primary criterion for identification of flagship hospitals was an explicit designation by the parent health systems on their websites, in the systems' official documents, or in press releases or through major media reports. In the few cases in which parent systems did not designate their flagships, we searched reliable online sources such as major newspapers and hospital reviews to see if there was an agreement among sources on the flagship status. If we could not unambiguously identify a flagship hospital in a multihospital system using these methods, the system was not included in the study. A health system could have more than one flagship hospital.

Because the concept of "flagship" often involves a role as a referral center for complex cases in a regional area small enough to have referrals from hospital to hospital within the same system, we excluded multistate national health systems (eg, Catholic Health Initiatives, Community Health Systems, Inc.) and health systems with no major teaching hospitals or no flagship(s) identified by the systems themselves. Non-acute care and stand-alone hospitals, hospitals with missing CMS Certification Numbers (CCNs) or unmatched CCNs or hospital types across different data files, and hospitals without a star rating, were excluded.

Our analyses were performed at both hospital and health system levels. In the hospital-level analysis, we grouped hospitals into "1-2 star," "3 star," and "4-5 star" rating categories. We first compared star ratings of flagship hospitals with those of major teaching hospitals in general (ie, hospitals in the CMS Hospital Compare database with resident to bed ratios ≥0.25 that were not designated as system flagship hospitals). We then compared the average flagship hospital and average nonflagship hospital star ratings pooled across all the health systems. To explore hospital-level characteristics that might be associated with flagship hospitals' performance on star ratings, we compared hospitals' teaching intensity, bed size, charity care,

and disproportionate share hospital (DSH) patient percentage between flagship and major teaching hospitals and between flagship and nonflagship hospitals. Differences were tested using two-sample *t* test with equal variances. We also compared hospital characteristics among hospitals with 1-2 stars, 3 stars, and 4-5 stars with use of one-way analysis of variance (ANOVA) with Bonferroni adjustment for multiple comparisons.

In the system-level analysis, we examined flagship hospitals' star ratings relative to the star ratings for other member hospitals in the same system. We assigned health systems to the following three groups according to their flagship hospitals' star ratings in comparison to other hospitals within their own systems: health systems in which flagship hospitals were rated the lowest among all member hospitals, health systems in which flagship hospitals were rated neither highest nor lowest or all hospitals within the system had the same star rating, and health systems in which flagship hospitals were rated the highest among all member hospitals. We compared system-level characteristics of the three groups. We calculated the average differences in uncompensated care payment, resident to bed ratio, DSH patient percentage, and total beds between flagship hospitals and nonflagship hospitals of the same health systems, and we also compared the differences across the three health system groups defined previously. We conducted an analysis of covariance (ANCOVA) to take system-level factors into consideration, including system size (total number of acute care hospitals in the system), systemwide teaching intensity, and systemwide charity care. The Bonferroni correction was used to adjust for potential problems of multiple comparisons.

Finally, to compare the diversity of star ratings within health systems and the diversity of star ratings nationwide, we generated a set of 100 pseudo systems each comprising six member hospitals (corresponding to the average number of member hospitals per "true" health system included in the study) that were randomly selected from all hospitals excluded from this study. We calculated and compared the average standard deviations of star ratings between the true health systems and this set of pseudo systems. Differences were tested using two-sample *t* test with equal variances.

Data management and statistical analyses were conducted using Stata SE, version 13.0 (StataCorp LLC, College Station, Texas).

RESULTS

Our final analysis included 599 hospitals in 113 health systems; 119 hospitals were flagships (four health systems each had two flagship hospitals, and one health system had three flagship hospitals). All other hospitals (n = 480) were designated as nonflaghips. On average, each health system had 6 member hospitals with star ratings, with a range from 2 to 22.

Flagship hospitals did have higher average star ratings than major teaching hospitals (mean star rating, 2.8 vs 2.3, respectively; *P* < .01; Figure). A larger proportion of flagship hospitals received four or five stars than did major teaching hospitals (29% vs 20%, respectively), and a smaller proportion of them received one or two stars (44% vs 59%, respectively; *P* < .05).
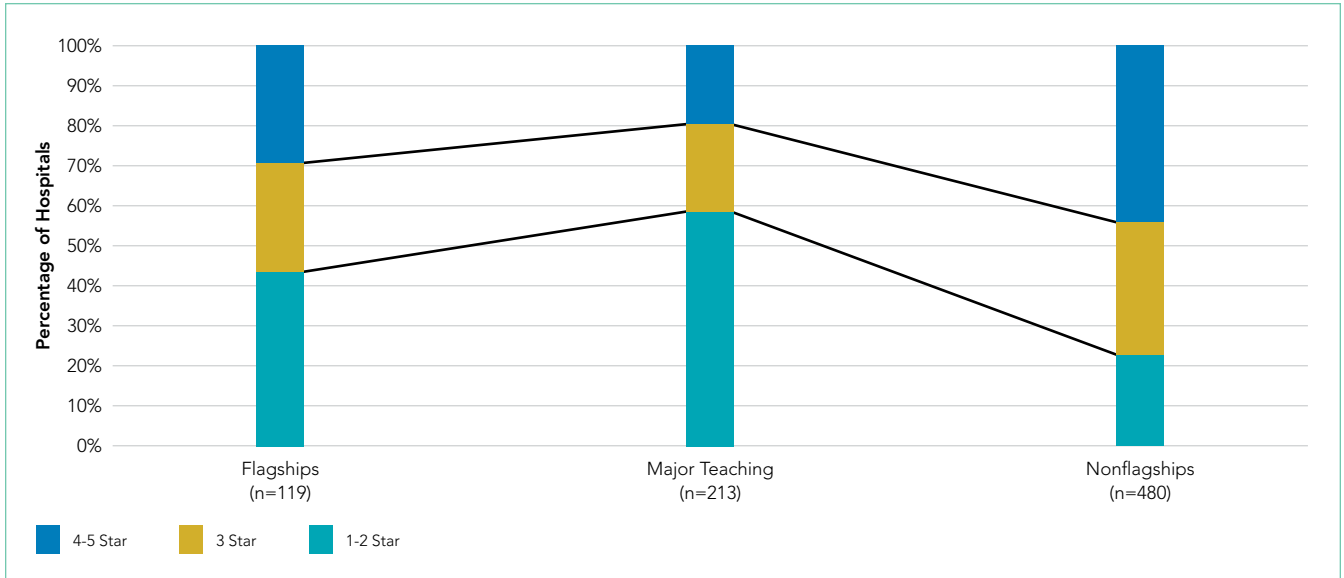
FIG. Percentages of flagship hospitals, major teaching hospitals, and nonflagship hospitals by their star ratings

TABLE. **Characteristics of Flagship Hospitals, Major Teaching Hospitals, and Nonflagship Hospitals by Star Rating Categories**

Hospital Characteristics	Flagships				Major teaching hospitals				Nonflagships			
	1-2 Star	3 Star	4-5 Star		1-2 Star	3 Star	4-5 Star		1-2 Star	3 Star	4-5 Star	
	(n=52)	(n=32)	(n=35)	*P* value	(n=125)	(n=46)	(n=42)	*P* value	(n=112)	(n=157)	(n=211)	*P* value
Resident to bed ratio (teaching intensity)	0.51	0.43	0.39	<.1	0.53	0.46	0.39	<.01	0.17	0.07	0.04	<.001
Total beds (#)	598	647	609	1.000	355	283	312	.1110	227	173	176	<.01
DSH patient percentage (%)	43	35	29	<.001	55	36	31	<.001	37	29	23	<.001
Uncompensated care payment per claim ($)	1,904	1,053	694	<.05	3,273	1,351	1,159	<.01	1,400	624	488	<.001

Flagship hospitals had lower star ratings on average, across all systems, than did nonflagship hospitals (mean star rating, 2.8 vs 3.3, respectively; *P* < .001). A smaller proportion of flagships received four or five stars than did nonflagships (29% vs 44%, respectively), and a larger proportion of them received one or two stars (44% vs 23%, respectively; *P* < .001).

As expected, flagship hospitals had significantly higher teaching intensity, larger bed size, higher DSH patient percentage, and higher value of uncompensated care payments than did nonflagship hospitals (*P* < .001 for all). On average, flagship hospitals were significantly larger but had lower DSH patient percentage and lower value of uncompensated care payments than did major teaching hospitals in general (*P* < .01 for all). In all types of hospitals, four- or five-star hospitals consistently had significantly lower DSH patient percentage (*P* < .001) and lower value of uncompensated care payment per claim (*P* < .05) than did other hospitals (Table).

In half of all health systems (n = 56), flagship hospitals were rated the lowest of all hospitals within that system; in approxi-

mately 20% of all health systems (n = 22), flagship hospitals were rated the highest. Flagship hospitals were more likely to have the lowest star rating in the system if the within-system difference in DSH patient percentage between flagship and nonflagship hospitals was relatively large. Within-system DSH patient percentage differences between flagship and nonflagship hospitals were 12.4%, 5.4%, and 3.5% in "flagship rated lowest," "flagship rated middle," and "flagship rated highest" systems, respectively (*P* < .05).

Average standardized deviations of star ratings for the 113 true health systems and 100 randomly generated pseudo health systems were 0.86 and 0.97, respectively (*P* < .05).

DISCUSSION

System-designated flagship hospitals did not generally have higher star ratings than did the other, smaller, community hospitals, either on average or within their own systems. In fact, the most common pattern observed was the system-designated flagship hospitals had the lowest star rating in their system.

Flagship hospitals in multihospital systems were, however, rated higher than major teaching hospitals in general. The safety-net role of many of the system flagship hospitals, as captured by relative DSH percentage, was the most important determinant of low star ratings. A high bed number and teaching status were not as strongly associated with low star ratings.

It is already well established that the CMS star rating system does not correspond to other global hospital ratings systems like those of *US News & World Report*, Healthgrades, or the Leapfrog Group.[11] Each global rating system uses a unique set of measures and weighting systems for those measures, so discrepancies among these systems are inevitable. Multihospital systems may feel that the positive reputation for tertiary care excellence held by a flagship hospital is captured in a rating system like US News that has an explicit reputation component[12] and that the US News rankings are more prominent in the public eye than are those of CMS. To the extent that the CMS star ratings do become more widely used by the public or by payers to establish narrow provider networks, the relatively low ratings of multisystem flagship hospitals may become a cause for concern for those hospitals and systems.

System-designated flagship hospitals are typically large teaching hospitals with higher levels of technology, more highly specialized services and medical staff, more extensive research programs and active clinical trials programs, and the ability to treat cases that are difficult or complex or instances of rare conditions. They are not generally, as it turns out, the hospitals in a given system that the CMS star rating system identifies as "best." In a number of multihospital systems, the system name is derived from the name of the flagship hospital (eg, Yale New Haven Health System and Montefiore Health System), which suggests that the system finds a marketing or branding advantage in being publicly identified with the name and positive reputation of the flagship hospital. Flagship hospitals may be designated as such because they have other attributes that patients, the community, and the system value, which may not be represented by the CMS quality metrics summarized by star ratings.

We did find a somewhat lower level of variation in star ratings in actual multihospital systems than in a set of randomly created "pseudo systems," suggesting the presence of some mechanism for quality management in those systems leading to a more similar set of star ratings than one would find in hospitals selected at random.

Our study has a few limitations. First, we excluded multihospital health systems without any major teaching member hospital, which was based on our observation that they do not usually designate their flagship hospitals or they do not have any identifiable flagship hospitals. There may be a small number of such health systems that have designated their flagship hospitals and were excluded from the study, but we do not believe it will change our key findings. Second, it was possible that multiple hospitals in the same health system reported under the same CCN (multicampuses will often use the flagship facility's IDs for the purposes of claims processing or cost and measure reporting), and therefore, the star ratings for the flagship hospitals reflected the performance of both the flagship hospital and the other member hospitals sharing the same CCN. We cannot fix the underlying reporting issue, and as a result, part of our analysis was probably more of a comparison of the "financial" flagship with other more loosely associated hospitals in the system. We could have in fact overestimated the flagships' star rating performance by including data of other better performing nonflagship hospitals.

CONCLUSION

System-designated flagship hospitals tended to have lower CMS Hospital Compare overall hospital quality star ratings than did nonflagship hospitals in the same multihospital systems. The characteristics of hospitals identified as system flagships do not seem well aligned with those associated with better performance in the star rating system.

References

1. Centers for Medicare & Medicaid Services. CMS updates website to compare hospital quality. December 21, 2017. Accessed October 28, 2019. https://www.cms.gov/newsroom/press-releases/cms-updates-website-compare-hospital-quality

2. Centers for Medicare & Medicaid Services. CMS Updates Consumer Resources For Comparing Hospital Quality. February 28, 2019. Accessed October 28, 2019. https://www.cms.gov/newsroom/press-releases/cms-updates-consumer-resources-comparing-hospital-quality

3. DeLancey JO, Softcheck J, Chung JW, Barnard C, Dahlke AR, Bilimoria KY. Associations between hospital characteristics, measure reporting, and the Centers for Medicare & Medicaid Services overall hospital quality star ratings. *JAMA.* 2017;317(19):2015-2017. https://doi.org/10.1001/jama.2017.3148

4. Wan W, Liang CJ, Duszak R, Lee CI. Impact of teaching intensity and sociodemographic characteristics on CMS hospital compare quality ratings. *J Gen Intern Med.* 2018;33(8):1221-1223. https://doi.org/10.1007/s11606-018-4442-6

5. Bilimoria KY, Barnard C. The new CMS hospital quality star ratings: the stars are not aligned. *JAMA.* 2016;316(17):1761-1762. https://doi.org/10.1001/jama.2016.13679

6. Chatterjee P, Maddox KJ. Patterns of performance and improvement in US Medicare's hospital star ratings, 2016–2017. *BMJ Qual Saf.* 2019;28(6):486-494. https://doi.org/10.1136/bmjqs-2018-008384

7. Chung JW, Dahlke AR, Barnard C, DeLancey JO, Merkow RP, Bilimoria KY. The Centers for Medicare and Medicaid Services hospital ratings: pitfalls of grading on a single curve. *Health Aff (Millwood).* 2019;38(9):1523-1529. https://doi.org/10.1377/hlthaff.2018.05345

8. Castellucci M. CMS star ratings disproportionately benefit specialty hospitals, data show. *Modern Healthcare.* 2018. Accessed October 28, 2019. http://www.modernhealthcare.com/article/20180314/NEWS/180319952

9. Joynt KE, Jha AK. Characteristics of hospitals receiving penalties under the Hospital Readmissions Reduction Program. *JAMA.* 2013;309(4):342-343. https://doi.org/10.1001/jama.2012.94856

10. Agency for Healthcare Research and Quality. Compendium of U.S. Health Systems, 2016. 2019. Accessed October 28, 2019. https://www.ahrq.gov/chsp/data-resources/compendium.html

11. Bilimoria KY, Birkmeyer JD, Burstin H, et al. Rating the raters: an evaluation of publicly reported hospital quality rating systems. *NEJM Catalyst.* August 14, 2019. Accessed February 19, 2020. https://catalyst.nejm.org/evaluation-hospital-quality-rating-systems/

12. Olmstead MG, Powell R, Murphy J, Bell D, Morley M, Stanley M. Methodology U.S. News & World Report 2019-20 Best Hospitals: Specialty Rankings. 2019. Accessed February 20, 2020. https://media.beam.usnews.com/8c/7b/6e1535d141bb9329e23413577d99/190709-bh-methodology-report-2019.pdf