Rating the Raters — Strengths and Weaknesses Assessment of the Four Public Hospital Quality Rating Systems

The comments in the table below reflect the discussion that the Rating-the-Raters group had about each rating system. These comments for each rating system were provided to the leaders of that rating system to solicit feedback.

Scientific Acceptability

(CMS) Hospital Compare Overall Star Ratings	Healthgrades Top Hospitals	USNWR (U.S. News & World Report) Best Hospitals	Leapfrog Hospital Safety Grade and Leapfrog Top Hospitals
Pro			
• Incorporates process, outcomes, and Patient Experience measures • The star rating is typically updated twice/year	• Does not use NHSN measures • Data updated annually • Volume-based weighting of performance	• Incorporates structure, process, outcomes, reputation • Compares hospitals in different ways to allow grouping of similar types of hospitals • Eliminated most PSIs • Eliminated all NHSN measures • Structural measures (e.g., volume) mitigate some of the measurement issues with outcome metrics • Reputation measure offers some information where there is a lack of more granular measures capturing the same concept • Risk adjustment for socioeconomic status in 12 mortality rates • Excluded external transfers from outcomes measures • Adjusts volume in each specialty to account for regional differences in Medicare Advantage Enrollment	• Incorporates process, structural, outcomes, and Patient Experience measures • Scientifically rigorous composite methodology • Impact Score for weighting approach is available and equally applied based on expert evaluation of impact, opportunity, and evidence base of measure • Z-scores are used to standardize data from measures with different scales and are applied to measures available for each hospital • Calculator available to hospitals to replicate measures scores, weights, and total score • The Safety Grade rating is updated twice/year • Hospitals must complete all applicable sections of survey to be included in Top Hospitals • Top Hospitals ratings separates peer groups (General, Rural, Teaching)
Con			
• Some concerns regarding measure weighting approach • Attempts to measure diverse hospital types together (major issue) • Concerns regarding approach for assigning stars (e.g., k-means clustering) • Use of PSIs, particularly PSI-90 and PSI-4 • PSI-90 and readmission weighted too heavily • Use of NHSN infection measures • No inclusion of clinical registry measures • Inclusion relatively unimportant imaging measures • No robust data audit process • Data lag can be up to 3 years from collection to release • Administrative data and NHSN data are not rigorously audited	• Not balanced measurement. No process measures or patient experience; composite relies only on outcome measures. Patient experience scores not included — shown as a separate rating. • Measures vary by state availability • Evaluating all hospital types together is a major issue • Proprietary methodology is not transparent and cannot be replicated and thoroughly evaluated • Assigning best hospitals based on percentiles of raw scores is a major concern (arbitrary threshold, top 5%) • No statistical testing done in "overall ranking" —just top 5% of hospitals • Inclusion of PSIs, particularly PSI-3, 4, 7, 12, 13 • Outcome measures and clinical cohorts (e.g., small bowel obstruction) are not conditions where patients typically can use the comparisons to decide where to go for care • No registry data included • Some complications are not related or relevant to the procedure (e.g., Legionnaire's disease is a complication for AAA repair) • Administrative data are not rigorously audited	• Still uses some flawed PSIs • Patient experience measures/Patient Experience used only for Procedures and Conditions rankings • Rankings done on hospitals of different types together (e.g., critical access, teaching) • Some concerns in their "reputation" measurement methodology with respect to sampling and ranking own institution highly (i.e., gaming) • Some specialties are ranked on reputation alone without any other measures of quality (Ophthalmology, Psychiatry, Rehabilitation, Rheumatology) • Limited use of registry data, except to give credit for participation in certain registries (but missing other major registries) • Hospitals missing patient safety data are assigned median patient safety score (PSI) for all hospitals • Administrative data are not rigorously audited • Risk adjustment different for procedures/conditions • Some metrics developed in-house have not been scientifically vetted in a robust fashion	• Voluntary, self-reported survey data account for 35% of score, and the survey data are not rigorously validated. This is a major concern. • Concerns about handling of hospitals not reporting (~50%) and about the corresponding missing data • Evaluating all hospital types together is a major issue • Very few clinically meaningful and rigorous outcomes data (exclusion of mortality is a major issue) • Inclusion of PSIs • Some particularly weak component measures (e.g., CPOE) • Hospitals still included if they don't answer Leapfrog survey • Methodology report does not comment on how reliable the primary and secondary data sources are with one another (used when primary not reported by hospital). • Use of legacy NQF Safe Practices (not maintained anymore), not intended to be scored, but rather intended to be used as an improvement tool • Administrative data are not rigorously audited • Use of NHSN infection measures • NHSN data are not rigorously audited • Leapfrog audits sample a very small number of hospitals annually • Audit results not released publicly • Voluntary, self-reported survey data account for 100% of score for Top Hospitals, but the survey data are not rigorously validated or relevant. This is a major concern. • Use of non-risk adjusted infection measures

Source: The Authors. Details of the rating process are available at www.RatingTheRaters.org