National Quality Forum Hospital Quality Star Rating Summit

ISSUE BRIEF

NOVEMBER 6, 2019

NATIONAL QUALITY FORUM

CONTENTS

FOREWORD

The National Quality Forum (NQF) was founded with the goal of improving care for every person, and from our founding, transparent measurement has been recognized as a powerful lever to advance patient-centered care and empower consumers. Our mission and vision rely on collaboration to address priorities common to diverse stakeholders. Bringing all voices together at the same time to discuss key issues is central to driving measurable health improvements.

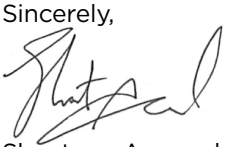
The Hospital Star Rating System was introduced by the Centers for Medicare & Medicaid Services (CMS) to summarize quality data in a way that is easy for patients and consumers to make informed healthcare decisions. Yet, reactions to the program are mixed and stakeholders on all sides raise important questions. Uncertainty surrounding the transparency and validity of the Star Rating program potentially creates confusion for patients and caregivers relying on the ratings to choose the best quality care—and concern among providers being rated and using the ratings for quality improvement.

NQF convened the Hospital Quality Star Rating Summit with more than 20 diverse experts from across the healthcare community to explore the opportunities and challenges that exist within the Star Rating program. We brought hospitals, payers, purchasers, and consumers together in one place to have an open and honest dialogue with the common goal of ensuring Hospital Star Ratings are accurate and actionable. The Star Rating Summit continues NQF's commitment to explore how measure sets and measurement systems can advance performance measurement objectives while increasing transparency and minimizing burden.

This issue brief was developed by the comprehensive deliberations inspired at the Summit and provides recommendations and considerations to guide future actions on the Star Rating program. Our recommendations focus on aligning and clarifying goals, increasing transparency, and enhancing the presentation of data for people using the Star Rating program. Implementing the recommendations will help ensure consumers, patients, and caregivers can easily assess quality of care information across healthcare providers and engage as meaningful partners in care decisions. It is important to note that NQF has not taken a position or provided any specific recommendations on policy changes to the Star Ratings program that CMS announced earlier this year; rather, the objective of this work is to guide future program development.

At NQF, everyone has an equal voice in driving healthcare improvements that provide the greatest value to all. Consumers and providers continue to need accurate and actionable quality information to make informed healthcare decisions, and NQF looks forward to championing this important discussion.

Sincerely,



Shantanu Agrawal, MD, MPhil

President and CEO

EXECUTIVE SUMMARY

The Centers for Medicare and Medicaid Services (CMS) designed the Hospital Star Rating Program to provide healthcare consumers with accessible quality performance data to inform healthcare decisions and help consumers select healthcare providers. A variety of stakeholders use and monitor this program, including hospitals, purchasers, and others beyond those originally intended. While some stakeholders find that the data support their needs, others have expressed concerns and identified several opportunities for improving the transparency, fairness, and usefulness of the program.

On August 23, 2019, the National Quality Forum (NQF) convened a multistakeholder Expert Panel from across the healthcare measurement enterprise to participate in the Hospital Quality Star Rating Summit to represent divergent perspectives. Summit participants represented the array of stakeholder perspectives: patients, consumers, employers and purchasers, healthcare providers, statisticians, and health plans. These participants shared ideas to strengthen the CMS Overall Hospital Quality Star Ratings system. NQF convened this Panel at a critical time for the program when recommendations from diverse perspectives can help drive improvements as CMS considers opportunities to improve the program. NQF self-funded this activity, without any external support.

Based on the Panel's earnest deliberations, demonstrated expertise in measurement science, and stakeholder priorities, NQF elicited three recommendations to guide actions for the Star Rating program.

NQF Recommendations:

1. **Be clear about the program intent and goals.** The program's methodology and design elements should align with the underlying intent and user needs.
2. **Be transparent about what the Star Ratings do and do not convey.** The Star Ratings are designed to provide a summary of measures on Hospital Compare to support consumer understanding. Clear communication about the interpretation of the program and its methods is imperative.
3. **Design data presentation to meet consumer priorities and other user needs.** There is an opportunity to enhance the presentation of the Star Ratings and user interaction with the summary data to make the program more actionable and relevant.

The Summit participants identified several key considerations for CMS to explore and test for integration in the next iteration of the Star Ratings program. These considerations align with the overarching themes that surfaced during the discussion.

Key Considerations from the Multistakeholder Panel:

1. More closely align the construction of the program and design decisions with the underlying intent of the program.
2. Consider regrouping measures to reflect clinically meaningful domains and service lines.
3. Expand the consideration of measures included in the program beyond Hospital Compare.
4. Consider alternative, simpler approaches to group scoring that improve understandability and transparency of measure weighting.

5. Consider an explicit approach to determine Star Rating thresholds to enhance the predictability and actionability of the summary data.
6. Consider eligibility criteria for hospitals to achieve a 5-star rating that reflect whether reporting measures in each domain has been achieved, particularly safety and mortality measures.
7. Balance the summary rating with the ability to drill down for more detailed information.
8. Differentiate methodology from the user interface and enhance the user interface.
9. Consider approaches to peer grouping that would be most meaningful to patients, for example, location, hospital characteristics, or service lines.
10. Expand the data sources to include a more comprehensive representation of patient populations served across hospitals.
11. Consider aligning program design elements across programs when possible.

INTRODUCTION

The National Quality Forum (NQF) has identified a need for greater consistency, transparency, and opportunity for multistakeholder input on how elements of measurement systems interact to assess performance. The comprehensive relationship of grouped measures and other programmatic elements offers significant potential to assess quality beyond a measure-by-measure model. With deep experience in convening and measurement science, NQF was poised to convene the Hospital Quality Star Rating Summit, which aligns with NQF's priorities to advance the conversation and science on measure sets and measurement systems and how to evaluate them.

On August 23, 2019, NQF convened a multistakeholder Panel of experts from across the healthcare measurement enterprise to participate in the Hospital Quality Star Rating Summit. Summit participants represented the spectrum of stakeholder perspectives: patients, consumers, employers and purchasers, healthcare providers, statisticians, and health plans. These participants worked together to develop key considerations to strengthen the Star Ratings system by exploring concerns and opportunities for improvement for the **CMS Overall Hospital Quality Star Rating system**. NQF convened this Summit at an important juncture: CMS is collecting input from stakeholders as it considers possible changes to the Hospital Star Ratings in 2021, following a rulemaking process in 2020[1]. The Summit participants discussed the key strategic issues in the design and implementation of the program and provided potential solutions to inform future testing by CMS.

CMS designed the Hospital Star Rating Program to provide healthcare consumers with accessible quality performance data to inform healthcare decisions on selecting hospitals. However, the data reported in the program are also used for additional purposes by more stakeholders than originally intended. For example, purchasers use the ratings to inform decision making for their members. While some stakeholders find that the data support their needs, others express concern and have identified several opportunities for improving the transparency, fairness, and usefulness of the program.

CMS has offered multiple opportunities to the public and key stakeholder groups to comment on the program via comment periods and national provider calls. Through this process, CMS has received over 800 comments which they have reviewed extensively. These comments were

also shared with the NQF Panel as background material for the Summit. In addition to soliciting feedback from users, CMS has convened multiple stakeholder-specific expert panels and workgroups to provide input and guidance on the program. Although summaries of these meetings are made public, the convenings and deliberations are not completely transparent. These convenings have included technical expert panels that provide input on methodology and maintenance of the program, patient and patient advocate workgroups focusing on priorities and usability of the program, and a provider leadership workgroup of hospital leaders and associations to address program implications and opportunities for improvement. Even with these opportunities to provide feedback, some stakeholders continue to express concerns regarding the program and the lack of clarity and communication on how and if CMS plans to address the issues raised. The CMS-convened activities also do not enable the three multistakeholder groups to collaborate within the same convenings.

This NQF-funded summit provided CMS a unique opportunity to hear all voices collaborate on the ongoing concerns raised by some stakeholders and presented an opportunity for NQF to serve as convener and facilitator of stakeholders, including CMS. With the guidance of a planning committee, NQF identified five critical topics based on the major themes across the comments and concerns addressed by stakeholders, including CMS, related to program design and implementation for the basis of the Summit agenda:

- Goals and intent of the program
- Measure selection and grouping
- Scoring methodology
- Accounting for patient risk
- Novel approaches

Although formal consensus on these complex topics was not sought, participants were able to find a common understanding and even agreement on some topics. The Panel identified 11 actionable considerations aligned with key themes for CMS to explore and test for integration in the next iteration of the Star Ratings program. NQF's three recommendations seek to provide high-level direction to CMS on where to focus its efforts on redesigning the program. Importantly, this issue brief is not intended to impact decisions regarding current reporting practices, payment policy, or any programmatic plans already announced by CMS. This brief summarizes the themes and key considerations that emerged during the Panel's deliberations and issues for further exploration by CMS.

APPROACH

NQF identified critical topics related to program design and implementation by examining prior public comments from key stakeholders, referencing related published literature, CMS's public documentation released on the program, and using input from CMS representatives. NQF used the measurement systems framework, developed in collaboration with Harvard Medical School in 2018, to guide its approach to the agenda and discussion. The 26 Summit participants were selected to represent a broad mix of stakeholders from the membership of NQF; a subset of the Panel representing each stakeholder group was selected to serve as a planning committee for the Summit. The planning committee worked closely with NQF staff to guide the agenda development and identify strategic issues for consideration. CMS and representatives from Yale's Center for Outcomes Research and Evaluation (CORE)—a CMS contractor for designing and maintaining the Star Rating Program—were invited to attend the Summit to participate in the discussion and provide the Panel with relevant background information on the program, its methodology, and updates under consideration. All Summit participants are listed in **Appendix A**.

The key considerations put forth here were identified based on input from various Panel members. They represent issues where there was careful consideration and significant deliberation from the Panel during the meeting. Recognizing that there was no voting process to indicate consensus and that the Panel represented a diverse set of stakeholders and perspectives, each key consideration is discussed in the context of those perspectives. NQF's recommendations seek to serve as additional framework to guide CMS's improvement efforts for the program. This brief will be released for NQF member commenting. Comments submitted by the NQF membership will be appended to the brief and shared with CMS and the public.

NQF RECOMMENDATIONS

NQF identified three actionable recommendations based on the Panel's deliberations regarding opportunities to improve the program, expertise in measurement science, and understanding of stakeholder priorities.

1. **Be clear about the program intent and goals.** The program's methodology and design elements should align with the underlying intent and user needs.
2. **Be transparent about what the Star Ratings do and do not convey.** The Star Ratings are designed to provide a summary of measures on Hospital Compare to support consumer understanding. Clear communication about the interpretation of the program and its methods is imperative.
3. **Design data presentation to meet consumer priorities and other user needs.** There is an opportunity to enhance the presentation of the Star Ratings and user interaction with the summary data to make the program more actionable and relevant.

THEMES AND KEY CONSIDERATIONS FROM THE MULTISTAKEHOLDER PANEL

The Star Rating Program uses a multistep process to select and group measures, score hospitals, generate a summary score, calculate Star Ratings, and apply reporting thresholds. Within these steps, there are multiple statistical methods and decision criteria that guide the handling of the data. Due to the step-wise approach to the program's methodology, many of the concepts discussed are closely related and intertwined. Eight cross-cutting themes emerged during the Panel's discussion of the program approach and methodology and served as the basis for the key considerations.

There is no available core set of measures that can be used to assess overall hospital quality, but we should continue to measure quality on the domains that are meaningful to patients and consumers.

The program intent is the foundation for all other program design decisions. The intent as initially conceived by CMS was to summarize the measures on **Hospital Compare** in a consumer-friendly display and interface; this summary is reported as the Star Rating. Both CMS representatives and the Panel members acknowledged that while this was the original intent, it has evolved into a tool used to indicate overall hospital quality used by stakeholders beyond patients and consumers. Several Panel members noted that CMS communicates that consumers could use the Star Ratings for hospital selection based on overall quality; however, as currently constructed, it should not serve that function. At least one Panel member noted that performance varies within hospitals as well as from hospital to hospital and that trying to draw inferences about quality across a broad array of hospital services from a relatively small number of measures centered on a few conditions may be misleading, especially for patients seeking care for conditions not included in the measurement.

KEY CONSIDERATION:

More closely align the construction of the program and design decisions with the underlying intent of the program.

Conceptually, the Panel supported a consumer-facing tool that allows an assessment of the quality of care provided by hospitals. Some stakeholders expressed support for the program's current methods to determine overall hospital Star Rating, noting quality is defined by the users and they find the information valuable. However, measuring overall quality and accounting for the complexity and breath of services offered by hospitals present challenges. Several participants questioned whether the current program captures the construct of "overall hospital quality" as the program's name implies. There remains no consensus among stakeholders or science to support a core set of measures that indicates overall hospital quality. Panel members encouraged CMS to clearly communicate the quality construct that it seeks to measure through the program. A coherent quality construct is important to guide the approach to measure selection, grouping, and weighting. As CMS re-examines the program, special attention should be given to each element of program design to ensure that it ultimately aligns with the stated program intent.

KEY CONSIDERATION:

Consider regrouping measures to reflect clinically meaningful domains and service lines.

Currently, there are seven measure groups, or domains, in the program: mortality (7 measures), safety of care (8 measures) readmissions (9 measures), patient experience (10 measures), effectiveness of care (11 measures), timeliness of care (7 measures), and efficient use of medical imaging (5 measures).

Measures are grouped based on similarities in the themes of the measure concepts. Measures in the mortality group, for example, include AMI mortality, COPD mortality, and mortality after surgery in one domain. These measures have little in common except their numerator statements, and in fact, the measures capture very different patient populations on different clinical care pathways, cared for by different provider specialty groups within the hospital. Some Panel members suggested that the current approach to measure grouping is not clinically coherent and has the potential to mischaracterize hospital performance for consumers. Further, consumers often seek information that is specific to their individualized healthcare needs, and a group of measures based on a domain that does not include a measure specific to their condition mislead.

Many Panel members supported the idea of reconstituting the groupings of measures into more clinically coherent measure groups. For example, a "cardiovascular care" group could include measures of AMI mortality, AMI readmission, and aspirin use after MI. If such an alternative grouping approach were used, summary performance of these groups could also be developed into service-line based ratings, which could supplement an overall rating and provide users additional information about the specific care they need.

The current Star Rating approach is an aggregation of measures available on the Hospital Compare website. The Panel questioned whether this approach provides the right performance data to consumers, purchasers, and providers.

To populate the Star Ratings program, CMS uses exclusion criteria for selecting measures from Hospital Compare which consider factors such as measure type (e.g., structural, nondirectional), number of hospitals publicly reporting, and reporting requirements for other programs, e.g., Inpatient Quality Reporting (IQR). CMS has also sought inputs from its technical expert panels to help inform its measure selection approach. The Panel also discussed several additional factors that should be considered to ensure that the measures address the needs of the program and that each individual measure's contribution to the overall score is fully understood.

These factors may include:

- Relevance and usability of the measures to consumers and patients
- Relationship (e.g., correlation) of the measure to other measures in the program
- Alignment with program intent and the other measures in the program
- Locus of control of the accountable entity (i.e., the hospital)
- Testing at the hospital level of analysis
- Accounting for social risk
- Balance of burden to implement the measure and the potential to improve quality
- Clinical relevance of the measure and the quality construct it is intended to capture

While this list is not exhaustive, it can serve as a guide for selecting a measure for the program and understanding its impact on the system. The Panel discussed the use of the **CMS Meaningful Measures framework** as a guide for helping to ensure that the program contains a parsimonious and relevant set of measures. CMS has primarily used this framework to remove measures from the program; however, consideration should be given to how it can be used to include relevant measures and fill gaps as well. The Panel presented perspectives that weighed both the burden of

expanding the set to be more comprehensive and inclusive of various clinical services, and the challenge of limiting the set to a small number of core measures. Some Panel members noted that if more clinically relevant measures were added to the set, that they should be measures that can drive improvement for hospitals and align with existing quality measurement efforts. Others pointed out that including more measures also raises the risk of burden and potentially provides more information for consumers to interpret.

KEY CONSIDERATION:

Expand the consideration of measures included in the program beyond Hospital Compare.

CMS acknowledges that it is currently constrained in its ability to include measures outside of Hospital Compare due regulatory requirements. However, many of the Panel members agreed that the current program and consumers would benefit from expanding beyond a summary of the available measures on CMS Hospital Compare to a more comprehensive representation of the specialties and services provided by hospitals. Given the complexity and breadth of services delivered in hospitals, a more comprehensive set of measures is needed to examine the services and markers of quality that are important to consumers and patients. While the Star Ratings program does not have its own statutory authority, the selection of measures into the program does rely on the public reporting and statutory authority for the underlying program of the measures, e.g., Inpatient Quality Reporting (IQR). One strategy for expanding the measures available on Hospital Compare could be to add measures to the IQR program with the stated purpose of using them in public reporting and not value-based purchasing.

Determining whether social risk has been accounted for in a measure is an important consideration in measure selection, as the Star Ratings program does not currently employ social risk adjustment at the measure group (i.e., domain) or star level. This type of risk adjustment is meant to account for social factors that may be out of the locus of control of the hospital, and may have a significant impact on patient outcomes. Currently, only adjustment for clinical risk factors is applied for certain measures within the program. Since there are no sociodemographic adjustments to the current measures used in Star Ratings, the Panel discussed the potential for social risk adjustment at the domain or overall Star Rating level. Members acknowledged that risk adjustment should be based on a conceptual and empirical relationship of social risk factors to the outcome(s) being measured. The Panel noted that applying risk adjustment at the domain level would be inappropriate for several of the domains since there is not a clear conceptual link between the measure and sociodemographic factors.

For example, risk adjusting process measures or measures in the safety domain for social factors would not be conceptually appropriate. The Panel focused its discussion of domain-level social risk adjustment on the readmissions domain and briefly considered whether overall adjustment at the Star Rating level would be appropriate. Some members noted that there were no scientifically proven methods for adjusting at the domain or Star Rating level. Others were concerned that the adjustment would equally apply to measures where there was no conceptual rationale for a difference in performance for sociodemographic factors. The Panel did generally agree that consideration of social risk adjustment at the measure level is appropriate and should be the basis of any consideration for selection of measures into the program and for any additional adjustment. Future work examining risk adjustment within measurement systems should consider the appropriateness, methodological challenges, and interplay of clinical and social risk adjustment the individual measure level and/or the measure group or domain level which may include multiple patient populations.

While there will be trade-offs if a simpler approach is used, the complexity of the methodology greatly impacts and even diminishes usability and provider efforts to improve quality.

The NQF Panel reviewed the latent variable modeling (LVM) approach used by CMS to estimate a group score for each of the measure groups. This approach uses a statistical model that assumes multiple measures reflect a single unobserved latent quality trait of a hospital that cannot be directly measured. The LVM approach attempts to measure this underlying quality trait through correlation and variation of measures in a given group.[2] CMS shared some advantages to LVM: It is data driven, accounts for the relationship between measures, and accounts for missing information as hospitals often do not publicly report on every measure. Conversely, CMS recognized that it is less intuitive, and some stakeholders have requested that CMS use a less complex approach.

KEY CONSIDERATION:

Consider alternative, simpler approaches to group scoring that improve understandability and transparency of measure weighting.

The Panel recognized that LVM is a common statistical methodology applied to data sets like the Star Ratings program; however, they suggested that simplicity may make it easier for both consumers to understand the ratings and for hospital to understand what they can do to improve. The LVM approach as applied assumes the existence of a single underlying trait shared by each measure in the group. Some Panel members challenged the validity of this assumption based on current measure groupings. There may be more than one latent variable driving quality in some cases. Even when analysis demonstrates that measures may not clearly share an underlying trait, they are still included in the group. In this case, measures that share or reflect the latent variable will be heavily weighted and drive performance in the domain.

Some concerns attributed to the LVM approach may be addressed by using a different approach to grouping. LVM is most effective when trying to capture a common pathway or process of care delivery underpinning the measures grouped together. The Panel encouraged CMS to consider grouping measures into clinical domains based on similar patient populations (e.g., patients with advanced cardiovascular disease) to better align related clinical processes and care pathways with the grouping approach. Panel members suggested that an explicit group scoring approach may offer similar results plus additional benefits, such as more predictability for hospitals to drive quality improvement for patients.

KEY CONSIDERATION:

Consider an explicit approach to determine Star Rating thresholds to enhance the predictability and actionability of the summary data.

CMS uses a method called K-means clustering to assign hospital summary scores to five Star Rating categories. This approach separates scores into categories such that hospital summary scores are most like scores in the same category and least like summary scores in other categories.[3] The method for K-means clustering is used to derive statistically based cut points. These cut points are for establishing the thresholds of performance for each star change across time periods based on the underlying measures and performance distributions across all hospitals in the program.

Many Panel members agreed that the approach to setting thresholds should be data-driven; however, some further suggested that CMS consider setting and freezing predefined performance thresholds for each star category in advance and for a period

of time to allow providers to have a line of sight to reaching the higher designation. Establishing predefined cut points would allow patients and the hospitals to understand how the star designation is achieved and how hospitals are trending over time. Hospitals may also benefit with predefined thresholds as a way to focus improvement activities in areas where their score will reflect those improvements. These thresholds would require updating when appropriate but should be data-driven and transparent.

While there was strong interest in enabling the viewing of performance trending over time, there was general agreement that incorporation of improvement into scoring should not be further explored. A hospital providing poor quality care should not receive a higher rating simply because improvement occurred if it is still providing low-quality care.

Simple methodology, when possible, supports understandability and interpretation of the rating. Methodologies like LVM and K-means clustering are important tools in data analytics but also cloud the interpretation of scores. When a simpler approach can be employed, it should be weighed against the other options and used when the impact of more complex options is less or equal.

Transparency to all users about how to interpret the Star Ratings and the program's methods is imperative; the risk of inappropriate use of the ratings is significant without it.

While the CMS Hospital Stars program is designed to convey information to patients, the summary data are also used by other stakeholder groups, such as payers, purchasers and employers, and the underlying data are used by hospitals for quality improvement. The Stars program helps these payers and purchasers select hospitals for inclusion in provider networks and provide rewards in incentive programs. A lack of clarity about the intent of the Star Ratings and its strengths and weaknesses can lead to misdirection for all stakeholders who use the program for their various purposes.

Although CMS shares detailed information about its methodology, many users do not have the context in which to understand and correctly interpret the meaning of the Star Ratings. This can lead to assumptions about Star Ratings representing an assessment of overall hospital quality rather than a summary of available measures—measures that were not selected with the intent of reflecting overall hospital quality. These issues may lead consumers to choose hospitals that were not the best at delivering the care they needed, or may cause payers to exclude hospitals from networks or disincentivize their use when they might have been the best place for a person with a particular condition to receive care.

Panel members pointed out several aspects of the program that could lead to misdirection of consumers based on the program's current construction and presentation:

- The aggregation of several hospitals under a single Medicare billing number makes it difficult for patients to assess the performance of the hospitals they are considering. The measures reported may be capturing the quality of care provided at another hospital in the system rather than the one at which the consumer seeks care.
- Care delivery can vary significantly from hospital to hospital, particularly for certain conditions and procedures, and this information and interpretation should be visible to users when they are deciding about their own care as well as to purchasers and plans when they are making network or payment decisions. This information cannot currently be discerned from the available measures.
- The lack of diversity of measures that make up the program leaves patients to potentially make hospital selections based on summary data that do not include measures relevant to them. For example, a 2-star hospital on

the website may also be the primary practice location for a surgery in a particular specialty. The lack of detail on service lines and their associated quality could potentially misdirect consumers away from a hospital where they may receive high-quality care for their condition.

- The models are sensitive to outliers. For example, a few outlier patients may skew the results of a measure and subsequently the domain and summary scores. One measure over several others may also drive the domain score for a hospital based on the current weighting and scoring approach.
- The number of measures for which hospitals have sufficient data to have results reported impacts hospital star ratings. For example, it is not clear to users whether Star Ratings represent a summary of all measures in the program and which measures or domains may not have been reported due to insufficient data.

KEY CONSIDERATION:

Consider eligibility criteria for hospitals to achieve a 5-star rating that reflect whether reporting measures in each domain has been achieved, particularly safety and mortality measures.

Despite recognition of the data limitations CMS faces, the Panel encouraged CMS to make transparent whether Star Ratings are sensitive to the number of measures reported and available to make comprehensive inferences about hospital performance, particularly in the safety and mortality domains, which resonate more with consumers. The current Star Rating methodology enables hospitals that do not have enough data to publicly report certain measures to be eligible to achieve a 5-star rating.

The Panel members agreed that eligibility for a 5-star designation should require that hospitals have a sufficient number of patients and data to assess performance; a lack of data should not benefit a hospital's Star designation. CMS should consider whether hospitals that do not have sufficient data for reporting on measures in the safety of care or mortality domains should be eligible to receive 5 stars. The potential for misdirecting patients exists. The consumer may assume that a 5-star hospital scores highly on safety and mortality when, in fact, these data are not included. At the other end of the spectrum, a small number of outlier cases should not significantly impact the overall star designation. The Panel acknowledged that establishing criteria to achieve a 5-star rating should be balanced with the goal of enabling maximum participation of hospitals in the program.

There is a need for the ratings to consider both simplicity and specificity (i.e., granularity at the level of hospital services or clinical conditions) to improve the usefulness of the information provided.

KEY CONSIDERATION:

Balance the summary rating with the ability to drill down for more detailed information.

The Panel supported the simplicity that a Star Rating offers, but encouraged CMS to consider providing more clinically focused ratings to enable patient-specific drill downs into the measures. There was general support for service-line rating drill down options to allow stakeholders to better interpret how hospitals received their rating designation and better align with the nature of specialty-based clinical care that patients are looking for when selecting a hospital. While consumers and patients want to have a comprehensive understanding of the hospitals when making decisions about care and to be able

to focus on their specific condition or procedure, others prefer and support the simplicity of a single star rating they can trust. While many consumers may not need to understand the methodologies employed to arrive at the results reported, they want to know and trust that the data presented have been analyzed with scientific rigor.

Recognizing that the Hospital Compare and Star Ratings website may only be one input that a patient considers in making a decision, the Panel suggested CMS further examine the choices that patients are making when they are selecting a hospital including insurance, reputation, and wait times and where consumers gather this information. A more comprehensive view into consumer choices at the time of selecting a hospital would inform the design and content of the tool to make it competitive with similar consumer-facing resources in the marketplace and improve usability.

There is an opportunity to enhance the presentation of the ratings to make them more actionable and relevant by considering how users want to view and interact with the summary information provided.

The Panel extensively discussed the nuances of the program methodology but also pointed out the necessity to differentiate which enhancements to the program would require methodological or data support and which could be implemented by providing a more interactive user interface. For example, providing drill down capabilities to service-line ratings would require both a modification to the webpage programming and additional data analysis to support any additional ratings that would be visible through the drill down.

KEY CONSIDERATION:

Differentiate methodology from the user interface and enhance the user interface.

The Panel encouraged CMS to consider human-centered design and effective consumer choice models to ensure the usability of the Hospital Compare website, including the CMS Hospital Star Ratings. Plain language interactive and media-based educational tools can assist consumers and others in understanding what is being measured and why it is important, and consumer groups encouraged CMS to involve patients in any website redesign efforts.

User customization of the ratings—for example allowing Hospital Compare users to assign their own weights to measures or select which measures contribute to an overall score—is one approach that CMS may consider. This flexibility would allow users to prioritize factors most important to them, which may lead to greater data-driven selection of providers. However there was also some caution about these options as users may not know the reliability of the measures they are selecting.

KEY CONSIDERATION:

Consider approaches to peer grouping that would be most meaningful to patients, for example, location, hospital characteristics, or service lines.

Patient and consumer representatives on the Panel noted that patients are generally selecting hospitals within a defined geographic region for most care decisions. Other Panel members noted that Hospital Compare already has these types of sorting features available and that it is imperative to be able to compare quality nationally for performance improvement efforts. For more specialized care, all hospitals do not provide the same types and levels of services.

On the other hand, purchasers argued that these national comparisons are essential in making decisions on network design and in designing value-based purchasing programs. Hospital

representatives found the national comparisons helpful to quality improvement efforts. In these cases, understanding the performance of local hospitals relative to national performance can help drive strategic decisions. The Panel briefly weighed the pros and cons of peer grouping via user-customized settings or statistical methods that would establish comparison groups. Generally, the Panel agreed that consumer-generated peer groups should be the focus of future efforts rather than a more methodological approach which would require significant testing and vetting.

All challenges with the program cannot be mitigated with statistical methodologies, but may be resolved by other approaches.

KEY CONSIDERATION:

Expand the data sources to include a more comprehensive representation of patient populations served across hospitals.

Several of the measures currently in use in the Star Ratings program reflect performance only for patients with Medicare fee-for-service plans. This limits stakeholders' view of performance on those measures for those using other forms of insurance. With the growing popularity of Medicare Advantage, particularly in some areas of the country, a limited picture of quality results for Medicare patients, too. In the future, the Star Ratings program should strive to represent the full consumer and patient population served by a hospital rather than only the Medicare fee-for-service population. To the extent CMS continues to collect Medicare Advantage encounter data and can validate the data as complete and accurate for calculating quality measures, CMS should work towards including Medicare Advantage encounter data in its calculations of claims-based measures. Continued efforts should also be undertaken to align measures with private payers. In order to align with the next generation of measurement, CMS should consider a multistakeholder effort to identify and prioritize a set of e-specified measures to assess performance of U.S. hospitals.

Reflecting on longer term program considerations, the Panel expressed the need to think beyond the boundaries of the current program. If the program were to be designed from scratch for today's world, how would we best guide individuals to select hospitals? The increased emphasis on price transparency and reliance on third-party applications as well as the potential to couple cost and quality to inform decisions were discussed as major factors worth considering. The Panel also suggested that interoperability and electronic measurement, real-time data sharing, and social media and reputation analysis be considered in the design and display of rating systems.

While specific to the Hospital Star Ratings, recommendations and considerations for improving the program may also apply to other programs and rating systems.

The Panel generally supported greater alignment in the methods used across CMS programs. While Panel discussions were specific to the Hospital Star Ratings, recommendations may also apply to other programs and rating systems facing similar design decisions. Though the intent across programs varies, they contain the same supporting components, for example, how to group and weigh measures, which scoring methods to use, how to account for differences in patient risk and missing data. Panel members suggested that CMS should consider aligning the approach to clinical and social risk adjustment across the various Stars programs, where appropriate.

KEY CONSIDERATION:

Consider aligning program design elements across programs when possible.

NEXT STEPS

This convening was the first of its kind for NQF and an example of how a multistakeholder convening approach can advance the science and design of quality reporting programs. CMS will consider the recommendations and considerations as part of its ongoing program re-evaluation activities. The recommendations will also support the deliberations of the CMS-convened technical expert panel in fall of 2019. As signaled by CMS, the next refresh of the Star Ratings with major methodological updates is anticipated for release in 2021 after rulemaking in 2020.

NQF's Measure Sets and Measurement Systems Technical Expert Panel is developing a standardized method for assessing measurement systems that builds upon the Measurement Systems Framework. The process will include a multistakeholder review of the program's goals and intent, measure selection process, grouping and scoring methods, and risk-adjustment approach and will equip NQF to examine additional programs in the future.

REFERENCES

1 CMS announces upcoming enhancement of Overall Hospital Quality Star Ratings [press release]. CMS News and Media Group. August 19, 2019. https://www.cms.gov/newsroom/press-releases/cms-announces-upcoming-enhancement-overall-hospital-quality-star-ratings.

2 Rabe-Hesketh S, Skrondal A, and Pickles A. Generalized multilevel structural equation modeling. *Psychometrika*. 2004;69(2):167-190.

3 MacQueen J. Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics.* Berkeley, CA: University of California Press; 1967:281-297.

APPENDIX A:
Summit Participants

Technical Expert Panel	Title	Organization
Nancy Foster, AB (Co-chair, Planning Committee)	Vice President for Quality and Patient Safety Policy	American Hospital Association
Cristie Upshaw Travis, MSHHA (Co-chair, Planning Committee)	Chief Executive Officer	Memphis Business Group on Health
Lisa Freeman, BA (Planning Committee)	Executive Director	Connecticut Center for Patient Safety
Lindsey Galli, BA (Planning Committee)	Patient Family Advisor and Director of Education	Patient and Family Centered Care Partners
Danielle Lloyd, MPH (Planning Committee)	Senior Vice President, Private Market Innovations and Quality Initiatives	American's Health Insurance Plans
David Nerenz, PhD (Planning Committee)	Director Emeritus of the Center for Health Policy and Health Services Research	Henry Ford Health System
Jody Amodeo, BSN	Senior Consultant	Willis Tower Watson
Andy Amster, MSPH	Director of the Center for Healthcare Analytics	Kaiser Permanente
John Matthew Austin, PhD, MS	Assistant Professor of Anesthesiology and Critical Care Medicine	Johns Hopkins University, Armstrong Institute for Patient Safety and Quality
Leah Binder, MA, MGA	President and CEO	Leapfrog
John Bott, MBA, MSSW	Consultant	Self-employed
Stephanie Clouser, MS	Data Scientist	Kentuckiana Health Collaborative
Maryellen Guinan, JD	Senior Policy Analyst	America's Essential Hospitals
Ben Harder, BA	Managing Editor and Chief of Health Analysis	U.S. News and World Report
Bala Hota, MD	Vice President, Chief Analytics Officer	Rush University
Omar Lateef, DO	President and CEO	Rush University Medical Center
Wendy Marinkovich, BSN, MPH, RN	Executive Director, Provider Measurement Programs	Blue Cross Blue Shield
Greta Martin, ScM	Healthcare Data Scientist	U.S. News and World Report
Kristine Martin Anderson, MBA	Executive Vice President	Booz Allen Hamilton
Janis Orlowski, MD, MACP	Chief Healthcare Officer	Association of American Medical Colleges
Louise Probst, BSN, MBA	Executive Director	St. Louis Area Business Health Coalition
Clarke Ross, DPA, MA	Public Policy Director	American Association on Health and Disability
Claudia Salzberg, PhD, MSE	Vice President, Quality	Federation of American Hospitals
David Shahian, MD	Vice President for Quality	Massachusetts General Hospital

Centers for Medicare and Medicaid Services Representatives	Title	Organization
Joseph Clift, EdD, MPH, MS, PMP	Technical Advisor	Centers for Medicare and Medicaid Services
Michelle Schreiber, MD	Director of Quality Measurement and Value-Based Incentives Group	Centers for Medicare and Medicaid Services

Yale CORE Representative	Title	Organization
Arjun Venkatesh, MD, MBA, MHS	Assistant Professor, Department of Emergency Medicine	Center for Outcomes Research and Evaluation (CORE), Yale

National Quality Forum Staff	Title
Shantanu Agrawal, MD, MPhil	President and CEO
Elisa Munthali, MPH	Senior Vice President, Quality Measurement
Ashlie Wilbon, MS, MPH, FNP-C	Senior Director
Nicolette Mehas, PharmD	Director
Andrew Lyzenga, MPP	Senior Director
Janaki Panchal, MSPH	Project Manager
Jordan Hirsch, MHA	Project Analyst
Taroon Amin, PhD, MPH	Consultant

APPENDIX B:
Key Considerations for CMS to Improve the Hospital Star Rating Program

1. More closely align the construction of the program and design decisions with the underlying intent of the program.
2. Consider regrouping measures to reflect clinically meaningful domains and service lines.
3. Expand the consideration of measures included in the program beyond Hospital Compare.
4. Consider alternative, simpler approaches to group scoring that improve understandability and transparency of measure weighting.
5. Consider an explicit approach to determine Star Rating thresholds to enhance the predictability and actionability of the summary data.
6. Consider eligibility criteria for hospitals to achieve a 5-star rating that reflect whether reporting measures in each domain has been achieved, particularly safety and mortality measures.
7. Balance the summary rating with the ability to drill down for more detailed information.
8. Differentiate methodology from the user interface and enhance the user interface.
9. Consider approaches to peer grouping that would be most meaningful to patients, for example, location, hospital characteristics, or service lines.
10. Expand the data sources to include a more comprehensive representation of patient populations served across hospitals.
11. Consider aligning program design elements across programs when possible.

APPENDIX C:
NQF Member Comment

NQF released the draft issue brief for member comment from October 11 to October 24; NQF received one member comment.

Betty Chu, MD MBA
SVP, Associate Chief Clinical Officer/Chief Quality Officer
Henry Ford Health System

We congratulate NQF on the excellent work with the Star Summit and the draft report. We particularly appreciate the approach taken by the group to first deal with issues of what the Star Rating system actually is and can be used for before moving on to technical improvements in the system. It is important for consumers to know that the Star ratings cannot be used to make decisions about the best hospital for specific elective procedures, nor is it an overall rating of hospital quality. We commend the work group and the report or making these points clearly. We support the recommendations in the direction of greater simplicity and transparency in the methods used for group scoring and for assigning overall Star Ratings based on total scores (the current "k-means clustering" approach). We also support the suggestion of having absolute, pre-determined cut-points for Star ratings, both for measure groups and for the overall score. As it stands now, hospitals have no idea whether an improvement in a given measure will improve either the group score or the overall Star rating. The setting of absolute cut-points would solve that problem, and would greatly enhance the value of the rating system for hospitals as they prioritize their quality improvement efforts.