NATIONAL QUALITY FORUM
Driving measurable health improvements together

Risk Adjustment Stakeholder Feedback Memo

VERSION 2

APRIL 14, 2022

This document is funded by the Centers for Medicare & Medicaid Services under contract HHSM-500-2017-00060I – 75FCMC20F0001

https://www.qualityforum.org

Contents

Introduction

Background

In partnership with the Centers for Medicare & Medicaid Services (CMS), the National Quality Forum (NQF) convened a Technical Expert Panel (TEP) of diverse multistakeholder experts to oversee the development of the **Technical Guidance** report for social and functional status-related (hereafter referred to as *functional*) risk adjustment in quality measurement. This project identifies comprehensive recommendations and analyses that are agreed upon best practices, by experts, as preferred methods for risk adjustment models. Since health outcomes are often the result of clinical risk factors as well as social determinants of health (SDOH), risk adjustment models that adequately account for these risk factors could inform policy decisions that aim to improve health equity and reduce disparities in care.

During the base period of this project (i.e., a 15-month performance period from 6/15/2020 through 9/14/2021), NQF conducted an **environmental scan** to identify current uses of functional and social risk factors (SRFs) in measurement. The TEP provided input on the environmental scan using relevant measure information worksheets submitted as part of NQF's **Consensus Development Process (CDP)** to examine current approaches to risk adjustment methods. Further, NQF and the TEP developed Technical Guidance in the base period, which is rooted in evidence and expert and stakeholder input on emerging issues in social and functional risk adjustment and within quality measurement. The guidance describes risk adjustment approaches agreed upon, by experts, as best practices for NQF's recommended consensus standards for risk adjustment models. These standards apply to both outcome and cost/resource use performance measures and some process performance measures at any level of analysis (e.g., health plans, facilities, individual clinicians, and accountable care organizations [ACOs]). The deliberations of the TEP and the recommended consensus standards for risk adjustment models will be considered by NQF's Consensus Standards Approval Committee (CSAC) for potential updates to the NQF endorsement criteria.

Purpose

For the next phase of work (the option period for this project is 9/15/2021 through 12/14/2022), NQF sought to garner broader perspectives by socializing the Technical Guidance with various healthcare quality stakeholders through a series of CMS-convened meetings and NQF-convened focus groups. Insights gained from this expanded stakeholder feedback effort will include anecdotal experiences related to developing risk adjustment models or experiences with viewing performance information for healthcare decision making, potential disagreement with the consensus standards and recommendations outlined in the Technical Guidance, proposed alternative approaches to account for functional and SRFs within quality measurement, and views on risk adjustment policies and procedures that are not apparent in NQF submissions.

The CMS-convened meetings include quality measure stakeholders, such as measure developers and federal staff working on the following items: quality measurement, quality improvement programs, public reporting programs, or alternative payment models. The goal of these CMS-convened meetings is for healthcare stakeholders to provide feedback on the Technical Guidance developed during the base period.

In addition to engaging with CMS-convened meetings, NQF convened focus groups across various stakeholder categories to increase awareness of the recommendations for the Technical Guidance. In particular, NQF reached out to individuals with minority viewpoints (i.e., those who disagree with the Technical Guidance recommendations and/or standards) to elicit its rationale for the TEP's consideration. Furthermore, to support the **White House Executive Order** to advance racial equity and support underserved communities through the federal government, NQF recruited members of communities that have been historically underserved. These communities included, racial and ethnic minorities, individuals with disabilities, those who live or have resided in rural areas, and those otherwise adversely affected by persistent poverty or inequality. Feedback received will be considered by the TEP during the option period for potential enhancements or refinements to the Technical Guidance.

Methods

Between September 15, 2021, and March 11, 2022, NQF presented the Technical Guidance to two (2), web-based, CMS-convened groups and conducted six (6), two-hour, web-based focus groups. The two CMS-convened groups were the Measure and Instrument Development and Support (MIDS) C3 Forum and the Quality Measurement Technical Forum (QMTF). Presentations at the MIDS C3 Forum and the QMTF meetings were conducted virtually. The MIDS C3 Forum provides a platform for education and outreach activities targeting CMS-funded measure developers. The topics covered by the C3 Forum focus on CMS' priorities for quality measurement, best practices in measure methodology, and the latest assessment findings on program use of measures. The QMTF is a CMS-wide monthly meeting regularly attended by CMS staff to share the experience(s) of different CMS programs on quality measurement or improvement. During these two meetings, NQF presented findings from the initial base period work, including the environmental scan and elements of the Technical Guidance, namely the consensus standards and the conceptual model.

For the NQF-convened focus groups, NQF recruited individuals representing healthcare measurement stakeholder groups, including measure developers, patients and consumers, payers and purchasers, quality improvement program leadership (QIPL) from both the public and private sectors, healthcare providers, and members of NQF-convened groups (e.g., Scientific Methods Panel [SMP], Standing Committees, and the CSAC).

Each focus group consisted of no more than nine participants. Individuals were identified through recommendations from the TEP, CMS program staff, NQF staff, members of the Patient and Caregiver Engagement Advisory Group (PACE), or web searches for stakeholders with specific expertise or insight. NQF also recruited individuals who have significant experience in the areas of quality performance measurement and measurement science; value-based program design; and those providing, paying for, and receiving care. The stakeholder categories for these focus groups are listed in **Appendix A**.

Focus group convenings were recorded and transcribed to support the development of key themes and considerations for updating the Technical Guidance (not included with this memo). Due to the potential sensitive nature of the topic under discussion (i.e., risk adjustment of social and functional risk factors), NQF has redacted the names and affiliations of participants. The memo identifies the stakeholder categories (i.e., consumer, health plan, clinician, and health system/hospital) along with the key themes and considerations from each of these focus groups (**Appendix C**).

NQF developed and used an in-depth discussion guide to facilitate the focus group discussions. The discussion guide provides the goals for each focus group along with a series of topics for discussion and open-ended feedback. NQF used a targeted facilitation approach for each focus group to promote an open exchange of information and to elicit as much insight as possible from participants. Similar topics were discussed by each focus group. However, because each focus group comprised different categories of stakeholders, NQF customized some of the discussion topics to solicit feedback that was relevant to the specific stakeholder category.

In advance of each focus group meeting, NQF shared the Technical Guidance, goals, and relevant topics with focus group participants. Brief descriptions of the topics as they were presented and the respective page numbers within the Technical Guidance are listed in **Appendix B**. Topics also referenced the Technical Guidance or other relevant resources. For example, the Patient and Consumer focus group referenced the Medicare **Care Compare** tool to provide context to the topic of discussion.

During the meetings, NQF staff presented the topic and used a round-robin approach to solicit feedback from participants. Follow-up discussions were also used to elicit additional points or more specific feedback.

Key Themes and Considerations

Several themes emerged for each topic and across each of the focus groups (Table 1). Descriptions of these themes as well as key considerations for updating the Technical Guidance are listed below.

Improvements to the Conceptual Model

Overall, focus group participants agreed that a strong conceptual model is useful and needed. However, participants identified several aspects of the model and its use for further consideration. Depending on how the guidance is interpreted, participants from the NQF-convened and Measure Developer focus groups noted a measure developer could spend months preparing a rigorous conceptual model and analyzing each factor that goes into the model. They recommended additional guidance on what constitutes evidence to support functional and social risk factor inclusion (i.e., quality, quantity, and consistency of evidence) as well as on the mechanism/approach to continuously test/iterate on the conceptual model. Some participants from the QIPL focus group suggested that the conceptual model should focus on health equity. There were several comments made about risk-adjusting for race. QIPL participants noted that race should not be used as a risk adjustment variable, as it is not clear what the variable truly represents. The Payer and Purchaser focus group commented that the core principle in the guidance stating that race and ethnicity could be used as a proxy for unmeasured SRFs could perpetuate the misconception that social needs and social risks are connected to race. The Payer and Purchaser focus group emphasized that social needs and social risk are not inherent to an individual's race. Additionally, others noted that NQF should endorse measures for their specific uses represented in the conceptual model and consider whether the intended use of the measure should be based on the quality of the data. Several focus groups noted that for measures that are used for quality improvement purposes, risk adjustment should not be applied.

Expanding the Locus of Control

Locus of control (LOC) refers to the scope of actions that the conceptual model assumes the measured entity can take to influence the outcome. For this theme, several focus groups expressed the need for clarification of what is meant by LOC within the Technical Guidance. Attendees of the Provider focus group expressed that there is variability at each level of measurement (i.e., clinician versus facility versus health plan), and that needs to be taken into consideration. As an alternative to measuring at these varying levels of accountability, it was suggested that the focus should be more on shared accountability and thinking about how measurement can leverage the broader LOC. For example, health plans may assist providers in reducing all-cause readmissions by coordinating patients' post-discharge care. Additionally, there were differing viewpoints when considering whether quality measures can be used to expand a measured entity's assumptions about the scope of actions it can take to influence outcomes beyond what it has typically viewed as within its LOC. For instance, it was shared within the NQF-convened group that to preserve the validity and purpose of measurement, the assumed LOC should match the measured entity's currently delivered interventions. Therefore, the assumed LOC should not reflect an expanded view of what the entity can or should do to affect outcomes. However, in other focus groups, there was agreement that measurement plays a role in expanding expectations for the actions that measured entities should take to influence the measured outcome; they also agreed that there should be incentives to support this expansion. Stakeholders are likely to disagree about how aspirational each individual measure should be. However, the conceptual model will help illuminate assumptions about the LOC so that the issue can be debated as part of measure review.

Burden to the Developer

Most comments regarding this theme came from the Measure Developer and NQF-convened focus groups. Participants noted that some analytic requirements of the Technical Guidance are ambiguous and may be burdensome to developers; they also noted that further standardization would help mitigate this burden. Specifically, they noted several requirements could require significantly increased time and cost for measure development, if not clarified: (1) For the standard that requires developers to consider certain social and functional risk factors in the conceptual model, more guidance is needed for the justifications for including or not including risk factors because a developer may create a narrative for any situation; (2) For the standard that requires calibration to be conducted within the overall population and within relevant at-risk subpopulations, several focus group participants noted that without further specificity with respect to identifying the relevant subpopulations, developers may spend a significant amount of time determining multiple relevant subpopulations, which could be burdensome or ill advised; 3) For the stratification standard, requiring developers to think through every single scenario for stratification can be burdensome. Participants in the Measure Developer focus group expressed that it is a heavy ask of the developers to make suggestions about stratification because the usefulness of stratification can vary greatly, i.e., variation due to the measure focus, the measure use, etc. Therefore, more specificity is needed for this standard to facilitate consistency in identifying by which subpopulation(s) should be stratified.

Developers also noted that they should not have to review measure specifications in all contexts but should enumerate considerations for measure users. For instance, NQF could endorse measures for a

particular care setting based on analyses most relevant to that setting and guide developers on what testing would be needed for applying the measure within other specific contexts.

Providing Clarity

Several focus groups noted that the guidance is ambiguous or unattainable in certain areas, namely the requirements for the following items: (1) empirical testing of risk factors and how to incorporate it into decision making for the final model and (2) developers describing the bias that exists in the absence of a risk factor in a data source, which is not possible, as there is no way to describe a bias in the absence of a variable. They also noted that some definitions within the guidance are unclear (e.g., "meaningfully influence," "locus of control," and "intended use") and require further clarification to prevent misinterpretation. For example, the QIPL focus group noted that a distinction needs to be made between asking providers to intervene on the risk factors themselves (e.g., homelessness) and asking providers to intervene on the link between the social and functional risk factors and the outcomes being measured (e.g., ability for homeless patients to access post-acute care medications to mitigate unintended readmissions to the hospital).

Stratification

Overall, focus groups agreed with the premise that stratification can help to reveal disparities in the measured outcome. Stratification can be defined as computing performance scores separately for different strata or groupings of patients based on some characteristic(s) (i.e., estimating multiple performance scores for each healthcare unit, one for each patient stratum, rather than estimating one overall performance score). Focus group participants stated that NQF has tremendous opportunity to address stratification, provide specifications for stratification, and drive stratification through measure endorsement. However, they also noted some of its limitations. For example, depending on the unit of analysis, measured population, etc., stratification can create small number reliability problems. Participants suggested that developers may choose to mitigate this by lengthening the data collection period, but this has trade-offs for reporting, quality improvement, and actionability. When determining the specific stratum, this should depend on actions that providers can take to influence the measured outcome, which are unique for that stratum. Within the patient and consumer focus group, stratification was a major focus of discussion. Participants underscored its importance but said that it only has utility to patients if they can see the characteristics of the patients and communities that the accountable entity serves. The importance of this would be twofold, 1) to identify if the accountable entity provides care to patients like them and 2) to see how well the accountable entity is performing on the care they provide to those patients. For instance, performance scores do not show the quality of care provided to patients who are from the LGBTQ+ community, are drug users, etc. With respect to the consensus standard that focuses on stratification, focus groups indicated that the guidance seemed to imply that developers should risk-adjust and stratify on the same variable, which is not appropriate. Lastly, NQF's guidance should distinguish risk stratification from stratification. For instance, a developer can stratify by a subpopulation if the intent is to simply display differences in performance for that subpopulation. However, if a certain subpopulation is at a higher risk for a measured outcome, then stratifying the measure separately for this subpopulation is referred to as risk stratification.

Risk Factor Selection

Several focus groups commented on the standards related to risk factor identification and selection. Generally, the NQF-convened focus group noted that the developer should discuss how their testing sample compares to the overall population of interest to ensure the risk adjustment model will perform well in the population for which it is intended. Additionally, there was attention shown to the need for more standardization of data, including data on race, ethnicity, and language. For instance, participants in the Payer and Purchaser focus group expressed that for race, there are only six classifications, and this does not represent true patient race. Providers need support on capturing race and ethnicity data to facilitate more consistency (and thus standardization) across the healthcare system.

Keeping to the concerns of consistency, several focus groups stated that for the standard which requires developers to consider certain social and functional risk factors in the conceptual model, caution should be used with respect to dual eligibility, as it is not consistently defined across the nation due to state-level variability in Medicaid eligibility requirements. Therefore, its use as a risk adjustment variable could inadvertently introduce bias.

Focus on Health Equity

Several focus groups mentioned that the conceptual model should be revamped to focus on health equity, stating that an equity viewpoint can further improve the model's application. Furthermore, to address inequities, it is important to know whether tools and resources will be available for providers to address inequities. It was suggested in the Provider focus group that NQF can identify disparities-sensitive measures that can be used to help identify and reduce or eliminate disparities in care. Disparities-sensitive measures are those that serve to detect not only differences in quality across institutions or in relation to certain benchmarks, but also differences in performance measure outcomes among populations or social groupings (race, ethnicity, language, etc.). NQF has previously conducted work in this area by developing **criteria** to determine whether a quality measure would qualify as "disparities-sensitive." Stratified reporting by groups or categories (e.g., race, ethnicity, gender, rural, and urban), which is subject to adequate sample size and availability to demographic data, could be useful for both accountability and quality improvement. Hence, further work to refine guidance on stratification in the report could advance the use of disparities-sensitive measures.

The Future of Measurement

Several focus groups considered how best to align and progress NQF's standards for risk adjustment in the context of an evolving health system that aspires to provide more holistic care that can drive improvements in outcomes for all patients. Focus group participants expressed that there should be a glide path that transitions from initially risk-adjusting measures for social and functional risk factors currently viewed to be outside of providers' control to decreasing adjustment for these risk factors over time, as healthcare providers and payers learn how to better reduce the impact social and functional risk factors have on the measured outcomes. The glide path should consider risk adjustment for social and functional risk in measurement and potentially move to methods of accounting for social and functional risk in payment models as they mature. Payment models should identify ways to establish incentives that advance the health of particular populations, not necessarily at the measure level but at a program level. However, until payment models can be structured in a way that appropriate resources

are allocated to improving care for disadvantaged patients, risk adjustment of some measures will be necessary to ensure fair comparisons between differently resourced entities.

Table 1. Summary of Focus Group Feedback

Topic	Focus Group/CMS Meeting	Minority Viewpoint	Key Theme
Standard Risk Adjustment Framework	• Measure Developers • National Quality Forum (NQF)-Convened Groups • MIDS C3 • QMTF	• Factors that are not modifiable should not be classified as "risks" (e.g., race, ethnicity, gender). Instead, stratification should be used to examine these variables. • Stratification is essential for not masking disparities. • There are not many methods for empirically testing and demonstrating definitively that disparities **are** under the entity's influence (e.g., due to discrimination). It is empirically easier to cast doubt on the fact that the reasons for the disparity **are not** under the entity's influence. Therefore, the assumption should be that **not** risk-adjusting for SRFs is the default. Then, an empirical analysis of how the SRFs **are not** under the entity's control would be appropriate to test for inclusion in a model rather than exclusion from a model. Any guidance on this would be helpful. • NQF could approve measures agnostic to a particular care setting or program but then specify what testing would be needed to apply the measure within specific contexts rather than just narrowly approving measures for a particular program.	• Improvements to the conceptual model • Stratification • Burden to the developer • Risk factor selection • Providing clarity

Topic	Focus Group/CMS Meeting	Minority Viewpoint	Key Theme
Conceptualizing the Model	• Providers • NQF-Convened Groups • Measure Developers • Quality Improvement Program Leadership • Payer and Purchaser • MIDS C3 • QMTF	• Organizational capabilities to affect risk factors may complicate and overburden measure users if they are not standardized. • Revamp the conceptual model from an equity framework perspective. • There should be a glide path to performing risk adjustment. The glide path should consider adjustments for social and functional risk in measurement and potentially move to adjustment in payment models as they mature. • There need to be incentives in place to expand the provider's LOC. • Race should not be included in the risk model because it is not clear what the variable truly represents. Using race and ethnicity as a proxy for social risk could perpetuate the thinking that social needs and social risk are connected to race, and this is not the case. However, if it is true that race is connected to being socially disadvantaged, is it ever appropriate for the provider to be accountable for this?	• Improvements to the conceptual model • Burden to the developer • Providing clarity • Expanding the LOC • Focus on health equity

Topic	Focus Group/CMS Meeting	Minority Viewpoint	Key Theme
Intended Use	• Measure Developers • NQF-Convened Groups • Quality Improvement Program Leadership • Providers • Payer and Purchaser	• Many agreed that measures that are used for areas of public health importance should not be risk-adjusted (e.g., safety measures). • Measures that are used for quality improvement purposes should not be risk-adjusted, and NQF should work in this space more. • There are some measure types that should never be adjusted (e.g., process, structure). • If measures are to be designed for specific use cases, then more guidance on validity and reliability testing is needed for those different use cases. • The guidance should reflect the considerations of the second report to Congress from the Office of the Assistant Secretary for Planning and Evaluation (ASPE); language specific to types of measures is not appropriate for this type [social risk] of adjustment.	• Stratification • Providing clarity • Burden to the developer • Expanding the LOC
Level of Measurement (i.e., Locus of Control)	• Measure Developers • Quality Improvement Program Leadership • Providers • Payer and Purchaser	• Some patients do not fit the premise that area-level adjusters are appropriate. Any individual's characteristics may not match that of the area's characteristics based on averages. • The guidance should distinguish between asking providers to intervene between the risk factors themselves and asking providers to intervene on the link between the SRFs and the outcomes being measured. • is/are related to the core principles on page 9 of the guidance. For the principle related to race and ethnicity used as proxy for social risk, fix the wording, as this could perpetuate the thinking that social needs and social risk are connected to race, and this is not the case.	• Improvements to the conceptual model • Expanding the LOC • Providing clarity • The future of measurement

Topic	Focus Group/CMS Meeting	Minority Viewpoint	Key Theme
Identifying and Selecting Potential Data Sources and Variables	• NQF-Convened Groups	• The focus group did not agree with the TEP that analysis of bias is possible for unknown data. However, an explanation of attempts to explain the bias should be required. • Once a better indicator of income, or a better social deprivation index, or other good indicators of social risk have been identified, there may be fewer concerns about putting these variables into a risk adjustment model because we can more readily measure an effect size for better quality of care.	• Burden to the developer • Risk factor selection • Providing clarity
Considerations for Determining the Final Risk Adjustment Model	• Measure Developers • Quality Improvement Program Leadership • Providers • Payer and Purchaser	• Payments should be adjusted for social risk, which may require additional resources for the accountable entity to achieve the same outcomes. • There should be a glide path to performing risk adjustment. The glide path should consider adjustments for social and functional risk in measurement and potentially move to adjustment in payment models as they mature. • There is not enough discussion in this report about bias and racism. Also, consider whether it is true that race is connected to being socially disadvantaged; if so, is it ever appropriate for the provider to be accountable for this?	• Stratification • Burden to the developer • Providing clarity • Risk factor selection • Expanding the LOC • The future of measurement • Focus on health equity
Stratification	• NQF-Convened Groups	• Stratification is more appropriate than risk adjustment when trying to address health equity. • Risk adjustment is more appropriate for clinical factors than social and functional risk factors. • Stratification is an alternative to risk adjustment, but it is not without its unintended consequences and burdens.	• Stratification

Topic	Focus Group/CMS Meeting	Minority Viewpoint	Key Theme
Accounting for Social and/or Functional Challenges in Provider Performance Scores	• Patient and Consumer	• Risk adjustment may disincentivize providers from accepting patients with higher risk.	• Improvements to the conceptual model • Expanding the LOC • The future of quality measurement
Use of Quality Measure Information in Selection	• Patient and Consumer	• None. All were in agreement that stratification is important and has a role in performance score reporting. This is aligned with the Technical Guidance.	• Stratification
Social or Functional Adjustment and Measure Type	• Patient and Consumer	• Process measures should not be risk adjusted, but outcome measures should show conceptual relationships. The total cost of care is different because sometimes you want to spend more to provide better outcomes.	• Improvements to the conceptual model • Stratification

Next Steps

NQF seeks to advance measurement science by developing Technical Guidance for measure developers that defines comprehensive recommendations and analyses that are agreed upon best practices, by experts, as preferred methods for functional and social risk factor adjustment in the context of measure development. While developed with input from a multistakeholder TEP with expertise in measurement science and risk adjustment, gaps in the information inevitably exist, especially for stakeholders underrepresented in its initial development, or groups not as familiar with NQF's endorsement process. Therefore, to ensure that the Technical Guidance and its recommendations and standards are relevant and have utility to the broader quality measurement ecosystem, NQF socialized this work to a broad group of diverse stakeholders, which is inclusive of those from medically underserved communities and those with underrepresented viewpoints.

The findings of this expanded stakeholder engagement, including the overarching themes and considerations, will be shared with the TEP in advance of the web meeting on May 12, 2022. With input from the TEP, NQF will update the Technical Guidance based on the key findings of the stakeholder engagement activities summarized in this Stakeholder Feedback Memo. Additionally, the Stakeholder Feedback Memo will be included as an appendix within the updated Technical Guidance.

Appendix A: Focus Group Stakeholder Categories

Table 2. Focus Group Composition by Stakeholder Representation*

Focus Group	Health Systems	Clinicians	Consumer/Patient	Payers	Purchasers	QMRI+
Provider Focus Group (n=8)	6	3	0	0	0	0
National Quality Forum (NQF)-Convened Focus Group (n=6)	3	2	1	1	0	2
Measure Developer Focus Group (n=7)	1	1	0	0	0	7
Quality Improvement Program Leadership Focus Group (n=7)	1	4	0	3	1	2
Patient and Consumer Focus Group (n=7)	0	0	7	1	0	2
Payer/Purchaser Focus Group (n=9)	2	3	0	4	3	2

**Counts are not mutually exclusive across columns within each focus group. For instance, in the Provider group, representatives were clinicians and provided perspectives based on their role within a health system.*

+*QMRI: Quality Measurement, Research, and Improvement. This category consists of stakeholders who conduct research on healthcare quality measurement and reporting. This group also includes measure developers, methodologists, accrediting bodies, certification boards, health policy and quality centers, and data services (analysis and aggregation) providers.*

Appendix B: Focus Group Topics

Accounting for Social and/or Functional Challenges in Provider Performance Scores (pages 4 – 5)

Hospitals or doctors are often compared by using quality measures, but it is not that straightforward. For example, if we want to compare two doctors both treating diabetic patients, one taking care of older patients with diabetes and other conditions and the other doctor taking care of younger and healthier patients with diabetes, we take patient difference into account by using statistical adjustments to make the patient population more similar across doctors. This enables us to compare doctors' quality of care as if they were treating the same patients. This is a very common approach to addressing patients' demographic differences, such as age and gender, and medical conditions. Patients' social risks or functional risks, such as living situations, income, and family support, can also impact their outcomes.

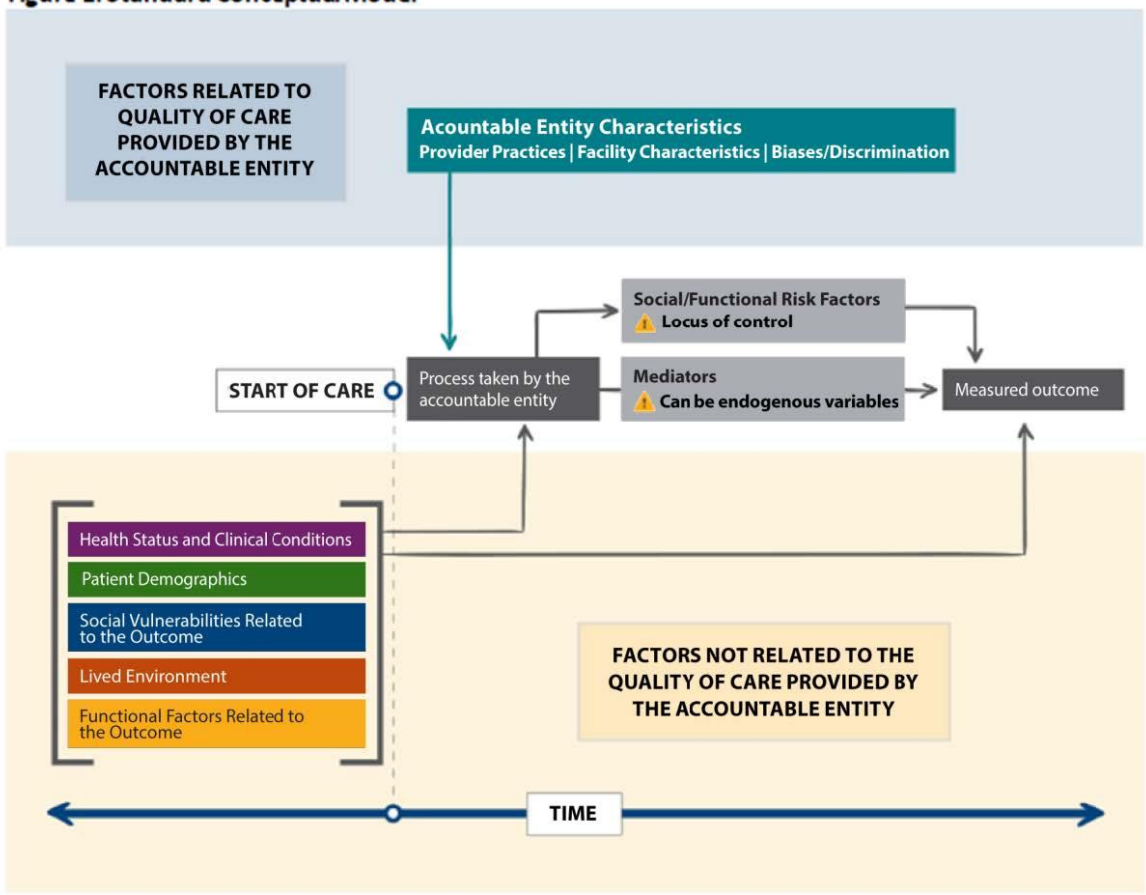
Standard Risk Adjustment Framework (pages 11 – 13)

The guidance identifies good and emerging best practices as minimum standards, supporting each of the steps in the process. These seven standards form a framework for the risk adjustment of health outcomes and offer guidance to achieving reliable and valid measure score results that can be compared across accountable entities. These minimum standards seek to consider limitations that measure developers may face. Often, developers must balance limited budgets as well as limited data availability and granularity with the analytic needs imposed by a detailed and complex conceptual model.

Conceptualizing the Model (pages 14 – 19)

A conceptual model visually describes the pathway between the social and/or functional risk factors, patient clinical factors, healthcare processes, and the measured healthcare outcome (Figure 1). By mapping these relationships, measure developers can begin to make clear and evidence-based decisions about the risk adjustment model. The Technical Guidance states that the pathway between risk factors and the care process should be illustrated and accompanied by evidence of the relationship. A well-developed conceptual model should be informed by clinical experts and patients, as well as clinical and population health research literature.

Figure 1: Standard Conceptual Model



Intended Use (page 19)

The Technical Guidance instructs that the specific intended use of the measure should be explained, to the extent known by the developer at the initial measure submission. The specific intended use of the measure may include public reporting; payment applications, such as value-based purchasing, shared savings programs, or other risk-bearing arrangements; quality improvement; or other policy and research applications. The intended use should be balanced with the LOC of the accountable entity to influence the social and/or functional risk factors identified in the conceptual model. A greater emphasis should be placed on the intended use for measures already in use and during the NQF endorsement maintenance process.

The TEP and other NQF-convened groups, such as the SMP, have noted that the evaluation of the appropriateness of a measure's intended use would be out of the purview of NQF endorsement. This type of measure evaluation would require different criteria depending on the intended use (i.e., evaluating validity and reliability for each use type). While the guidance acknowledges that the conceptual model should inform whether/how to adjust or stratify for social/functional risk in the context of the specific intended use, NQF does not currently endorse measures for specific intended uses. However, the Technical Guidance recommends that the conceptual model should outline the

evidence in the context of the LOC and the specific intended use of the measure. Moreover, developers should re-evaluate social and/or functional risk adjustment when adapting measures for other uses.

Level of Measurement (i.e., Locus of Control) (pages 18 – 19)

Within the conceptual model, it should be clear which steps and processes the accountable entities can influence to improve the measured outcome and those they cannot influence. Evidence to support these decisions can be from a combination of sources, such as expert opinions, literature reviews of peer-reviewed articles and white papers, and/or internal empirical analyses. Therefore, the conceptual model must consider the most appropriate and relevant level of measurement (e.g., ACO, health plan, and individual clinicians) during the development process.

The guidance instructs developers to consider whether social and/or functional risk factors confound the quality-outcome relationship. Specifically, what is the level of evidence that accountable entities can mitigate that impact of the social or functional risk factors to the outcome measured? Furthermore, the conceptual model should consider whether it is feasible for accountable entities targeted by the measure to diminish the impact of social or functional risk factors.

Factor Selection (pages 19 – 22)

Once social and/or functional risk factors are identified within the conceptual model, the guidance states that the developer should examine the data sources and variables available to capture these identified risk factors. The conceptual model will facilitate the selection of factors for risk adjustment. Although social and/or functional risk factors may be identified in the conceptual model, there may be data limitations that will have an impact on their use as variables within the risk model. The guidance acknowledges that there are data limitations for factor selection and states as a minimum standard that if social and/or functional status risk factors are not available but are included in the conceptual model, the developer should document this occurrence and provide a rationale explaining whether and how the omission of these data might bias the results.

Calibration (pages 23 – 25)

The guidance states as a minimum standard that risk adjustment model performance must be assessed in terms of calibration. Risk model calibration statistics inform whether the risk adjustment model-predicted probabilities are, on average, close to the average observed probabilities. To adequately assess the impact of social and/or functional risk, risk adjustment model calibration must be examined within at-risk subpopulations (e.g., racial categories). Moreover, these subpopulations should be defined in the conceptual model.

Additionally, all risk models should be tested and vetted to examine whether they are significantly under- or overpredict for important subgroups with social or functional risk. If a risk factor is not included in the model, the developer should, at a minimum, provide evidence that this does not bias the measure results for that group or subgroup. Developers should be transparent about their approach to and interpretation of the results.

Considerations for Determining the Final Risk Adjustment Model (pages 25 – 28)

The guidance states that social and/or functional risk adjustment may not be appropriate for all measures. The Technical Guidance recommends that measure developers should examine each measure on a case-by-case basis to determine the appropriateness for social and/or functional risk adjustment, taking a measure's conceptual relationship with individual risk factors into consideration.

The intent of this guidance is to provide a standard approach to social and/or functional risk adjustment within performance measurement. As such, the minimum standards outlined are to provide developers with the necessary tools needed for NQF endorsement, respective to social and/or functional risk adjustment. Although NQF does not control how measures are implemented or used, it is important to signal that program polices have an impact on accountable entities caring for populations with social and/or functional risk.

The federal government and other public payers must operate within the constraints of statutory requirements (e.g., budget neutrality, types of measures allowed for a program, and inclusion of rural or office-based physician practices [i.e., small numbers in measure denominators]). These constraints may impact decisions, often informed by legal guidance, about the final risk adjustment model.

Risk Stratification (page 25 – 26)

Risk stratification refers to the division of a population or resource services into distinct, independent strata or groups of similar data, thus enabling the analysis of the specific subgroups. This approach can be used to more clearly show the areas in which disparities exist or a need is present to expose the differences in results. Risk stratification is an important analysis to conduct in conjunction with risk adjustment to identify health and healthcare disparities.

The guidance states, as a minimum standard, that measure developers should demonstrate appropriate use of both risk adjustment and risk stratification, including providing rationale and strong evidence in cases in which the measure is not risk-adjusted or stratified. Developers should report stratification specifications (e.g., categories and combinations of SRFs) by specific and relevant subgroup categories, such as racial/ethnic categories, gender, socioeconomic status, and functional status. Additionally, stratification should be conducted to show within- and between-providers' performance by key subgroups to further determine which providers perform well or are poorly serving disadvantaged, or at-risk, populations.

Policy Considerations (pages 27 – 28)

In its most recent report to Congress, the Assistant Secretary for Planning and Evaluation (ASPE) concluded that resource use (e.g., admission/readmissions, cost of care) measures used in value-based purchasing programs should be adjusted for social risks, whereas many outcome measures should not. The rationale for this recommendation for resource use measures was as follows: Compared to accountable entities serving a more advantaged population, the accountable entity serving more socially at-risk individuals may require additional resources to achieve the same high quality care. Conversely, for outcome measures, ASPE asserts that the accountable entity has some control of the care given in the care setting. Thus, according to ASPE, outcome measures should not be adjusted for social risks.

Appendix D: Examples of Approaches to Social and/or Functional Risk Adjustment (pages 43 – 70)

For each section of the Technical Guidance, an example is provided within this appendix. The examples, which include figures, tables, and verbatim text, have been extracted from performance measures that have been evaluated by NQF's CDP, which are all NQF-endorsed. These measures were part of the illustrative set that was identified within the TEP-informed environmental scan.

Use of Quality Measure Information in Selection (Medicare.gov)

As an illustrative example, Medicare.gov displays information on how clinicians, hospitals, health plans, etc., perform on certain quality measures so that patients and consumers can see how well a provider provides healthcare to the patients they serve. Medicare beneficiaries can compare the performance of providers within their area by using the Care Compare search feature. We are not discussing the specific Care Compare tool but rather, how social and/functional risk factors can be accounted for in public displays of provider performance.

Social or Functional Adjustment and Measure Type

Some have argued that statistical adjustments to make the patient population more similar across providers should only apply to certain types of measures. Broadly, measures can be categorized into three groups: process of care, outcomes of care, and cost of care. For example, measures capturing a process of care could include diagnostic screening, outcomes can include change in a patient's blood pressure or blood sugar, and cost of care refers to the total resource utilization for a patient or patient condition over a year.

Appendix C: Focus Group Data Capture

Provider Focus Group

Topic Areas

- Policy Considerations (pages 27 – 28)
- Conceptualizing the Model (pages 14 – 19)
- Intended Use (page 19)
- Level of Measurement (i.e., Locus of Control) (pages 18 – 19)

Technical Guidance Sections

Considerations for Determining the Final Risk Adjustment Model

POLICY CONSIDERATIONS (PAGES 27 – 28)

Focus Group Notes (Key Considerations)

- We need to be cautious of how it is acceptable that higher social risk produces outcomes that are not as good – not acceptable
- Payments should be adjusted for social risk, which may require additional resources/inputs on accountable entity to achieve the same outcomes. This is widely consistent with how the Technical Guidance is framed.
- Adjusting for payments provides the necessary resources for upstream issues that are not reimbursed.
- There may be an unintended consequence if quality measures are risk-adjusted because it may mask disparity by adjusting and pull attention away from certain areas.
- We should risk-adjust cost/resource use measures but not outcome measures.
- There is discrimination within provider statistics, such as high no-show rates due to institutional stigma.
- Other alternatives to compare provider performance include days cash on hand, debt ratio, differentiator incentive to close disparities on the back end rather than performance, increase reimbursements or multiplier for clinicians, and **adjusting benchmarks to address health equity and vulnerable populations**.

Emerging Themes

- Expanding the LOC

Minority Viewpoint

- Payments should be adjusted for social risk, which may require additional resources/inputs on accountable entity to achieve the same outcomes.
- An alternative is adjusting benchmarks to address health equity and vulnerable populations.

Conceptualizing the Model

INTENDED USE (PAGE 19)

Focus Group Notes (Key Considerations)

- We need to better understand the populations we serve through the collection of proper data and understanding race/ethnicity to increase the quality of care we provide.
- Intended use should only be permitted if the user has engaged in agreed-upon risk adjustments.

- In order to improve the outcome of subpopulations, we need to first identify them (e.g., race, ethnicity, language, and gender identity).
- Intended use should be based on the quality of the data, not whether it is important or not.
- For the already well-established quality measures, stratify by race/ethnicity through an endorsed mechanism to look at the most disparities-sensitive measures to establish a health equity index.
- The outcomes are publicly reported for each health system and what the composite index is; therefore, it puts the outcomes into context.
- There needs to be a standard in place on how to measure these measures that touch on all domains.
- Measures are not the same across systems; therefore, we need to unify by eliminating fragmentation and have distinct definitions that move across entities in care.
- The concern is that depending on the feasibility on measure, it will be placed into a domain instead of being balanced with the LOC.

Emerging Themes

- Stratification
- Intended use of measures
- Cost index for quality improvement

Minority Viewpoint

- To improve the outcome of subpopulations, we need to first identify them (e.g., race, ethnicity, language, and gender identity).

LEVEL OF MEASUREMENT (I.E., LOCUS OF CONTROL) (PAGES 18 – 19)

Focus Group Notes (Key Considerations)

- Clinicians can do a better job if a small population has high social risk compared to their whole panel having social risk. Instead, look at the percent of their "load" when caring for high social risks.
- There is too much variability based on the accountable entity/individual. An attempt to make sure the system can support and accommodate those who cannot take off work to go to an appointment needs to be made.
    - Ex: A subset of patients wants Friday afternoon time slots, but clinicians are not available.
        - What about those who cannot take off work and can only come in on the weekend?
        - Perhaps a system accommodation is needed and not a clinician.
- There is variability at each level (clinician versus system vs population), and this needs to be taken into consideration yet also interconnecting.
- It is concerning that individual providers are penalized with readmissions rates when it is better suited for team-based care (i.e., ACO with resources) and connections with community resources in care delivery.
- Have one LOC that has different weights. Providers are limited by time, and if a provider has a patient population with higher risk factors, they may not seem as efficient compared to another provider who does not have as many patients with high risk. factors.

- The developer must consider social and functional risk factors because the LOC of a measure will be unique based on the particular circumstances of the market, the system, and the way value-based contracts are working.
- Err on the side of being more involved to have the LOC be as broad as possible.
- Allocation of resources should be dependent on the size of the organization/biggest need.
- Clinics do not have the same resources, and these ultimately drive the outcomes resulting in the importance of LOC.

Emerging Themes

- Expanding the LOC

Minority Viewpoint

- There is too much variability based on the accountable entity/individual. Attempt to make sure the system can support and accommodate those who cannot take off work to go to an appointment.

Overall

Focus Group Notes (Key Considerations)

- How measurement deals with systemically biased data gaps? While some are collected, some are also very significant and biased in a systemic way.
- There is worry about eligibility because it varies between states. We need to have conformity across the country.
- Has NQF given any consideration to intercommunity stakeholder engagement as the LOC as participants and accountability in quality measures without noting what those measures should be?
- Ensure not to build disadvantage into the guidance.
    - Having a home health aide
        - Many who have decreased functional status do not have a home health aide.
        - Do not disadvantage the disadvantaged based on their insurance.

NQF-Convened Focus Group

Topic Areas

- Standard Risk Adjustment Framework (pages 11 – 13)
- Conceptualizing the Model (pages 14 – 19)
- Intended Use (page 19)
- Risk Factor Selection (pages 19 – 22)
- Calibration (pages 23 – 25)
- Risk Stratification (page 25 – 26)

Technical Guidance Sections

Overview

STANDARD RISK ADJUSTMENT FRAMEWORK (PAGES 11 – 13)

Focus Group Notes (Key Considerations)

- General agreement with standards and its sequence, but a few improvements can be made on its details.
- More discussion on what should be included in every model is needed. For example, communication-related issues should be considered (i.e., primary language, cognitive status) because that will vary considerably across patient populations.
- Definitions of what data are and are not available need to be standardized.
- The problem that must be solved is whether there is variation in risk factors across all providers that must be accounted for. This should be an element of exploration for the conceptual model.
- Risk adjustment is a method for improving the validity of measures. We want to preserve between-provider differences while addressing the within-provider disparities. This is not burdensome; rather, it is essential to use the best-available science to demonstrate validity.
- Dual-eligibility criteria vary state by state because Medicaid eligibility varies state by state. This is a large issue with this particular data variable.

Emerging Themes

- Improvements to the conceptual model
- The validity of measures
- Providing clarity

Minority Viewpoint

- This group seemed to agree that the minimum standards are not burdensome as written and are worried that this concern as written in the Technical Guidance currently may inadvertently threaten the validity of measures.

Conceptualizing the Model

LEVEL OF MEASUREMENT (I.E., LOCUS OF CONTROL) (PAGES 18 – 19)

Focus Group Notes (Key Considerations)

- The conceptual model is vital, and the technical details flow from it.
- Designing measures to expand the LOC is not as important as preserving the validity of the measure to measure what it is supposed to be measuring.

- The developed model has to be based in the evidence but also flexible enough to address future needs and desires for healthcare. This is especially true as we move towards value-based payment models rather than the more typical model of care.
- Adjustment of patient-reported outcomes (PROs) may not be appropriate.
- Again, until we can define for the measure developers what is and is not within an entity's LOC, we will continually disagree about this decision for each individual measure. Uniform agreement is not likely, but it should not be the burden of the developers to define and defend this for each measure.
- It is not necessarily appropriate to put items into a measure to change accountability; rather, the measure should match the accountability level.
- This group did not accept the premise that expanding the LOC is appropriate for all measures. Measures should match, not expand, the agreed-upon scope of responsibility of the measured entity. This scope is expanded in some programs (e.g., arranging for transportation and receiving payment for that) and more classically defined in others (e.g., not compensated for transportation assistance). This is appropriate.

Emerging Themes

- Improvements to the conceptual model
- Expanding the accountable entity's LOC
    - The LOC should match the measured entity, not expand it, to preserve the validity and purpose of measurement.
- Providing clarity on intended use
- The future of measurement

Minority Viewpoint

- Some patients do not accept the premise that area-level adjusters are appropriate. Any individual may not match the area's preconceived characteristics based on averages.

INTENDED USE (PAGE 19)

Focus Group Notes (Key Considerations)

- Intended use is particularly important to understand for high-stakes applications, such as payment.
- Risk aversion is a concern due to the way measures are used in practice, and risk adjustment has implications for this matter.
- Measure developers should signal how the measure is going to be used, but the focus group also recognized that this may or may not in fact happen following endorsement.
- NQF should conduct/carry out/perform endorsement for use because of these concerns and the statistical sensitivity of measures, depending on their design. They may not be valid for certain types of uses.
- Implementation of the measure can be burdensome to the accountable entity. The use of the measure should outweigh the burden that it undergoes to use the measure. This may change depending on the use.
- Templates for different uses could be valuable for both the evaluation and development of measures.

Emerging Themes

- Providing clarity on intended use
- Burden
    - Reducing the burden through standardization
        - Of measure use
        - Of measure testing
        - Of measure evaluation
    - The current NQF process does not accommodate endorsement decisions regarding use well.

Minority Viewpoint

- None. The group was in agreement about the intended use.

Identifying and Selecting Potential Data Sources and Variables

FACTOR SELECTION (PAGES 19 – 22)

Focus Group Notes (Key Considerations)

- The conceptual model is a viable guide to what should be analyzed for inclusion in a model. If the literature says to look at something and you really have not tried, that does not sit well.
- However, it is not possible to analyze the impact of a variable for bias if you do not have the data available. Instead, developers should be asked to make their best effort to explain their efforts to capture the variable and their best efforts to mitigate bias and how they will improve this at maintenance.
- It has sometimes been surprising to hear about what measure developers do not have access to. It may be possible to ameliorate this issue through improvement to data access agreements.
- The problem with data is not in most developers' control; it is a problem at the site of care (e.g., Race, Ethnicity, and Language (REL) data are just not collected.)
- However, there is a ***data paradox:*** We hear that there is a humongous amount of data, but we also hear that there are still gaps within those data for particular items of interest.
- Including tips on how to improve data collection efforts in the future would be helpful.
- Dual-eligibility criteria vary state by state because Medicaid eligibility varies state by state. This is a large issue with this particular data variable.
- This must be assessed on a measure-by-measure basis to assess whether there is a fundamental flaw in the measure or whether this would merely improve an already valid measure.

Emerging Themes

- Dealing with the unknown
    - Data paradox (see above note)
- Burden
    - Reducing the burden through standardization
- The future of measurement

Minority Viewpoint

- The group did not agree with the TEP that an analysis of bias is possible for unknown data. However, an explanation of the attempts to explain the bias should be required.

Empirically Testing the Adequacy of the Risk Model

CALIBRATION (PAGES 23 – 25)

Focus Group Notes (Key Considerations)

- The focus group did not discuss this item.

Emerging Themes

- The focus group did not discuss this item.

Minority Viewpoint

- The focus group did not discuss this item.

Considerations for Determining the Final Risk Adjustment Model

STRATIFICATION (PAGES 25 – 26)

Focus Group Notes (Key Considerations)

- The group agreed with the premise that this __ can help to reveal disparities in the measured process/outcome.
- There is a concern that adjusting for social risk may harm patients with SRFs by not making potential differences in quality/performance transparent. If adjustment is made for differences between groups, the developer should also analyze the difference's root cause. Namely, is the difference due to variability in the population? Or is it due to concerns regarding equity or applications of the measure?
- It is a heavy ask for developers to make suggestions about stratification because the usefulness of stratification can vary greatly by local circumstances or an organizational level. A developer should not be required to think through numerous scenarios; instead, they should give broad guidance on the level of data they have, and measure users should determine this instead.
- Depending on the unit of analysis, measured population, and other factors, stratification can create small numbers problems. Developers may choose to mitigate this issue by lengthening the data collection period, but this has trade-offs for reporting and quality improvement.

Emerging Themes

- Stratification
    - Risk stratification should be distinguished from stratification.
- Burden
    - Elements of the technical guidance that suggest developers should think through every single scenario are burdensome.
    - Stratification can even be burdensome.
    - Developers should not have to review measure specifications in all contexts but should enumerate considerations for measure users.

Minority Viewpoint

- Stratification can be a solution, but it is not without its unintended consequences and burdens.

Overall

Focus Group Notes (Key Considerations)

- The document should be reviewed to add content on what could be added to address the future of the healthcare system.
- Generally, the developer should discuss how their testing sample compares to the overall population of interest to ensure the risk adjustment model is applicable outside of testing.

Measure Developers Focus Group

Topic Areas

- Conceptualizing the Model (pages 14 – 19 of the Technical Guidance)
- Standard Risk Adjustment Framework (pages 11 – 13 of the Technical Guidance)
- Intended Use (page 19)
- Considerations for Determining the Final Risk Adjustment Model (pages 25 – 28)
- Appendix D: Examples of Approaches to Social and/or Functional Risk Adjustment (pages 43 – 70)

Technical Guidance Sections

Conceptualizing the Model (pages 14 – 19 of the Technical Guidance)

Focus Group Notes (Key Considerations)

- Participants agreed that the conceptual model is useful and that measure developers should undergo this rigorous exercise.
- However, depending on how the Technical Guidance is interpreted, a measure developer could spend months preparing this and analyzing each factor that goes into the model. Some guidance on what constitutes evidence to support factor inclusion would be helpful.
- The ability to "meaningfully influence" a factor should be explained more.
- More explanation and clarification on the light grey boxes are needed because they do not appear to be at the start of care and are confusing.
- More explanation and clarification on what should and should not be included in the teal box of the Conceptual Model figure is needed.
- The Technical Guidance should include an explanation of the mechanism to continuously test/iterate on the conceptual model as you fill it in to continually monitor for unknowns.
- "The most important information is unknown and unknowable."

Emerging Themes

- Improvements to the conceptual model
- Burden
    - Elements of the technical guidance that ask for ambiguous analysis requirements are burdensome to developers.
    - Some definitions are unclear (e.g., "meaningfully influence") to developers.
- Providing clarity

Minority Viewpoint

- Organizational capabilities to affect risk factors may complicate and overburden measure users if they are not standardized.

Standard Risk Adjustment Framework (pages 11 – 13 of the Technical Guidance)

Focus Group Notes (Key Considerations)

- Developers overall sought more clarity on a few fine points in the minimum standards. The number of minimum standards referred to is noted in parentheses below.
- Additional guidance on Sexual Orientation and Gender Identity (SOGI) is important to clarify inevitable questions (2).

- Measures for pediatric populations may not have been considered when constructing this list (2).
- What will happen to a measure if a developer does not consider these factors (2)?
- It may be possible for a developer to come up with a narrative for any situation, whether including or not including a factor. If guidance can be set for these justifications, it may prevent this from happening (2).
- As in other groups, developers pointed out that dual eligibility is not a consistent population across the nation because of the variability in Medicaid eligibility requirements (2).
- There is no way to describe a bias in the absence of a variable (3). What type of bias is this referring to?
- The Technical Guidance was ambiguous about the requirements of the empirical testing of risk factors and how to incorporate them into decision making for the final model (4).
- Calibration can be thought of in multiple ways, and so, additional details may be needed. Many felt this standard would be burdensome or ill advised and that more specificity is needed (6).
- There is stratification, and then there is risk stratification—the Technical Guidance should not use these terms interchangeably (7, and generally).
- Differing opinions were shared about using stratification in concert with risk adjustment. One developer noted that using them in conjunction for the same variable is not appropriate. Others did not raise this concern, indicating they would not agree. One developer found this to be too burdensome and ambiguous (7).
- Creating community standards and ranking data elements that are regularly used for their quality may alleviate future arguments about them.

Emerging Themes

- Improvements to the conceptual model
- Stratification
    - Risk stratification should be distinguished from stratification.
- Burden
    - Reducing the burden through standardization.
- Risk factor selection
- Providing clarity
    - There is no way to describe bias in the absence of a variable.

Minority Viewpoint

- Factors that are not modifiable should not be classified as “risks” (e.g., REL, gender). Instead, stratification should be used to examine these variables.
- Stratification is essential for not masking disparities.
- There are not many methods for empirically testing and demonstrating definitively that disparities are indeed under the entity’s influence (e.g., due to discrimination). It is empirically easier to cast doubt on the fact that the reasons for the disparity are *not* under the entity’s influence. Therefore, the assumption should be that *not* risk-adjusting for SRFs is the default. Then, an empirical analysis of how the SRFs are *not* under the entity’s control would be appropriate to test for inclusion in a model rather than exclusion from a model. Any guidance on this would be helpful.

Intended Use (pages 19 of the Technical Guidance)

Focus Group Notes (Key Considerations)

- The group agreed that process and structural measures should not be risk-adjusted, and there are some outcome measures in which the risk factors will cancel each other out. However, different people may categorize the same measure in different ways.
- There is not a bright line for whether an entity can or cannot meaningfully influence a risk factor, so the final decision may have to incorporate the measure's use.
- There are certain levels of analysis (e.g., plan level) in which you do not need to level the playing field/are not comparing plans; as a result, risk adjustment is not valuable in these cases.
- Developers are not only choosing whether to adjust or not; they are also making simultaneous choices about stratification and exclusions as well. Risk adjustment should be used when a single numerical value to compare across entities is most valuable.
- Measure developers are and should be required to be attuned to the unintended consequences of measures.
- Measures being used for payment have taken the bulk of endorsement work, but measures are used for other purposes as well. The discussion of health equity is helping to balance these conversations more.

Emerging Themes

- Stratification
- Providing clarity on intended use
    - By measure type
    - By program
    - By level of analysis

Minority Viewpoint

- Many agreed that measures that are used for areas of public health importance should not be risk-adjusted.
- Measures that are used for quality improvement purposes should not be risk-adjusted, and NQF should work in this space more.
- There are some measure types that should never be adjusted (e.g., process, structure).

Considerations for Determining the Final Risk Adjustment Models (pages 25 – 28 of the Technical Guidance)

Focus Group Notes (Key Considerations)

- The group was asked specifically about whether risk adjustment should always be at the measure level, and there were several responses.
- Some noted that stratification can be more appropriate than risk adjustment when trying to address inequities.
- Others noted we should not ask too much of the measures; risk adjustment should be used to make comparisons across entities as fair as possible. This is not a solution to every problem, and measures should not be expected to solve all problems.
- Several agreed that measure developers should think carefully about the context in which measures are going to be used because this should affect the design of the measure. For

example, the Merit-Based Incentive Payment System (MIPS) has a complex payment bonus if an entity has a particularly complex patient panel.

Emerging Themes

- Stratification
- Providing clarity on intended use
- The validity of measures

Minority Viewpoint

- None – the group was in agreement about this topic.

Appendix D: Examples of Approaches to Social and/or Functional Risk Adjustment (pages 43– 70)

Focus Group Notes (Key Considerations)

- The group did not discuss this item.

Emerging Themes

- The group did not discuss this item.

Minority Viewpoint

- The group did not discuss this item.

Overall

Focus Group Notes (Key Considerations)

- The group was asked about their thoughts on the role of risk adjustment in health equity.
- Some noted again that risk adjustment has the potential to cause unintended consequences. There is a trade-off between fairness of comparisons and masking differences in performance.
- Once a measure is risk-adjusted, it takes away a measure user's ability to calculate their own score. This limits the measure's use as a QI tool or the user's ability to track and improve on their own performance.
- There may be legal restrictions or policy guidance against risk-adjusting for certain SRFs that tie measure developers' hands. The Technical Guidance should make this information explicit.

Quality Improvement Program Leadership Focus Group

Topic Areas

- Conceptualizing the Model (pages 14 – 19), specifically:
    - Intended Use (page 19)
    - Locus of Control (pages 18 – 19)
- Considerations for Determining the Final Risk Adjustment Model (pages 25 – 28)
- Overall

Technical Guidance Sections

Conceptualizing the Model (pages 14- 19)

Focus Group Notes (Key Considerations)

- There should be a glide path in place to performing risk adjustment. The glide path should consider adjustments for social and functional risk in measurement and potentially move to adjustment in payment models as they mature.
- Because CMS has more care coordination codes, there may be better outcomes available for patients with social and functional risks.
- There need to be incentives in place to expand the provider's LOC.
- Risk adjustment cannot be thought of as static because better providers can help patients overcome social and functional risk barriers, particularly if we arrive at a payment model that provides more resources for those in more vulnerable areas.
- It is frustrating for providers and facilities to be held accountable for something that is entirely outside of their control.
- The focus group had various perspectives on the use of social and functional risk adjustment at the measure level:
    - On one hand, the group did not entirely agree that social risk adjustment at the measure level is appropriate—specifically, the use of **race** because it is not clear what the variable truly represents in a risk model.
    - On the other hand, the group noted that measures at the individual level need to consider SDOH to ensure fair comparison across all facilities. It would be difficult to compare fairly without these adjustments.
- Focus group members noted that once a better indicator of income or a better social deprivation index has been identified, or other good indicators of SDOH have been identified, we may have fewer concerns about putting these variables into a risk adjustment model because we can more readily measure an effect size for better quality of care.
- We saw an analysis from a measure developer on the impact of dual-eligible beneficiaries; we saw only in the fifth quintile that there was essentially no overlap in the outcomes. Thus, there is an opportunity to improve and achieve similar outcomes for various patient populations.

Emerging Themes

- Risk factor selection
- Expanding the LOC

Focus Group Notes (Key Considerations) – March 10, 2022

- It is imperative for providers to meet the needs of the populations that they serve. This model shows that all needs can be addressed with the same process regardless of what the need is.
- Revamp the conceptual model from an equity framework perspective.
- Having an equity viewpoint in the conceptual model can further improve application.
- The framework (design) for adequate discharge planning should be reassessed, as some factors are outside of the control of the hospital (e.g., community resources).

Emerging Themes – March 10, 2022

- Expand the LOC
- Focus on health equity

Minority Viewpoint – March 10, 2022

- It is imperative for providers to meet the needs of the populations that they serve. This model shows that all needs can be addressed with the same process regardless of what the need is.
- Revamp the conceptual model from an equity framework perspective.

INTENDED USE (PAGES 19)

Focus Group Notes (Key Considerations) – March 10, 2022

- More guidance for Validity and Reliability for different use cases is needed.
- We are worried about the income effect, not the marginal incentive according to high-stakes measure use.
    - Below incentive = business case to bring into reward zone
    - It is challenging to get out of the penalty zone when caring for a higher-risk population.
        - We do not have to sacrifice penalty by risk-adjusting.
    - Deal with wealth effects through grants, and provide off-setting payments.
        - Only do not change incentives (equity approach).
- The guidance should distinguish between asking providers to intervene between the risk factors themselves and asking providers to intervene on the link between the SRFs and the outcomes being measured.
- There are ways to set up incentives to consider populations, not necessarily at the measure level but at a program level. All___ does not need to rely on the measure.
- Are vulnerabilities being assessed the same? If we are using the traditional model for tying care to incentives, it creates a concept of how many strata we can create.
- Additional support to address the additional data may create a certain level of expected outcomes in the end.
- Look into how risk stratification can be part of intended use and the purpose, such as the potential collaboration, and assess the impact of social needs.
- ____ is/are concerned that different risk adjustment models for programs are related to star ratings.
- How a measure should be defined is important when using intended use and determining which one to use when it comes to stratification and structuring approaches.
    - To address inequities, it is important to know tools will be used alongside the measure.
- The conceptual model is helpful, as the Indian Health Service (IHS) participates in many of the programs and has large health inequities. However, there is concern in rural areas due to fewer

hospitals in the area that can show a large spike (i.e., it affects the Star rating). Risk adjustment is important and takes these things into consideration.

Emerging Themes – March 10, 2022

- Stratification
- Expanding the LOC

Minority Viewpoint – March 10, 2022

- More guidance for Validity and Reliability for different use cases is needed.
- ___ is/are worried about the income effect, not the marginal incentive according to high-stakes measure use.
- The guidance should distinguish between asking providers to intervene between the risk factors themselves and asking them to intervene on the link between the SRFs and the outcomes being measured.

LOCUS OF CONTROL (PAGES 18-19)

Focus Group Notes (Key Considerations) – March 10, 2022

- The LOC focuses mainly on financial and labor resources to overcome any social barriers.
    - This can get "fuzzy" for small practices when looking at measurement on an individual level compared to an organization.
- Measurement adjustments could determine that more resources are needed for organizations caring for specific populations rather than the challenges faced by those populations.
- Regarding the discussion of whether intended use tries to expand the LOC to address social or functional risk factors:
    - it is reasonable to say that we are trying to expand the LOC and provide instruction and resources when expanding; and
    - we do not think that the current LOC should automatically be considered for risk adjustment.
- It is important to focus on the LOC in this conceptual model and what should be risk-adjusted versus not risk-adjusted.
- Connecting people to care through SDOH was focused through an evaluation of which metrics are most meaningful (e.g., reducing the homeless population).
- Some quality measures are applied to a broad group.
    - The LOC is very different when applied at a health plan versus a provider.
        - This raises the question of how we can leverage the broader LOC that a health plan has to help providers perform better.
- There are differences in screening SDOH, when that is screening being done, and at what level.
    - There are structural inequities outside of the hospital to address.
- Expand the LOC considering the facility, provider, and patient.
- Hold hospitals accountable for disparities in readmissions and try to incentivize to reduce the gap.

Emerging Themes – March 10, 2022

- Expanded LOC

Minority Viewpoint – March 10, 2022

- Whether intended use tries to expand the LOC:
    - it is reasonable to say that we are trying to expand the LOC and provide instruction and resources when expanding; and
    - we do not think that the current LOC should automatically be considered for risk adjustment.

Considerations for Determining the Final Risk Adjustment Model (pages 25 – 28)

Focus Group Notes (Key Considerations)

- The approach to risk adjustment is not a statutory requirement. There is no requirement on if/how we must perform risk adjustment.
- There is a lot of opportunity present with risk stratification. With a few exceptions, we do not provide, either publicly or confidentially, performance reports that are stratified in any way.
- We should ensure a glide path from adjusting measures to adjusting payment models since there are multiple levers that CMS, or any payer, can pull.
- The continuous process of looking at measurement and payment adjustment is a critical question to ensure that equity is at the forefront our thinking and that there are no structural barriers to improvement or measure outcome achievement. Process measures are under the control of the facility or provider and probably should not be risk-adjusted.
- Focus group participants were encouraged by the work of the TEP and the ability to look at the adjustment of measures in the same manner as the clinical adjustment of measures since the payment environment is a policy decision, and the quality measures may have to adapt to it.
- NQF has a tremendous opportunity to address stratification, provide specifications for stratification, and drive stratification through measure endorsement.

Emerging Themes

- Stratification
- The future of measurement

Focus Group Notes (Key Considerations) – March 10, 2022

- Risk adjustment goes with official value-based purchasing.
- It is imperative to keep in mind the small populations when risk-adjusting.
- Incentive design influences risk adjustment models, specifically looking at improvement versus attainment.
    - Yes, risk adjustment does matter, and how the measure is used as QI and payment reporting seems to have a gap.
- The context of how the measure will be used in the community that goes beyond the patient level and recognizes that neighborhoods have an independent impact on people's health and their outcomes.
- It was stated that it may be more about use-case than the measure itself. This can be addressed through a checklist from NQF that notes improving quality of care, presents inequities in care, and does not sacrifice criterion (e.g., validity, reliability, etc.). This checklist will demonstrate how to achieve these goals.

- The stratification depends on doing something unique for that stratum, such as an equity lens versus a quality lens.
- It is important to know what works for each group to create useful stratum by stating what things work and for whom does that work.
- Different statistical rules apply to different models.
- It is important to be cautious about stratification in small populations or small sample sizes.
- A large issue is the lack of reliable data in terms of race and ethnicity.
- Risk adjustment should still be at the measure level, but not all measures should be risk-adjusted.
- Stratification can vary based on how you stratify, such as the Stars program using peer-grouping based on the number of measures being reported.

Emerging Themes – March 10, 2022

- Stratification
- Caution with small populations when stratifying
- Expand the LOC

Minority Viewpoint – March 10, 2022

- It was stated that it may be more about use-case than the measure itself. This can be addressed through a checklist from NQF that notes improving quality of care, presents inequities in care, and does not sacrifice criterion (e.g., validity, reliability, etc.). This checklist will demonstrate how to achieve these goals.

Overall

Focus Group Notes (Key Considerations) – March 10, 2022

- Until payment models can be structured in a way that more resources are allocated to disadvantaged patients, risk adjustment is necessary to ensure fair comparisons between differently resourced entities.
- Stratification is the strongest tool we currently have in improving health equity through measurement. Requiring stratification by certain subpopulations may help.
- Payment decisions to improve health equity should be kept separate from decisions about specific quality measures.

Special Populations Focus Group

Topic Areas

- Accounting for Social and/or Functional Challenges in Provider Performance Scores (pages 4 – 5)
- Use of Quality Measure Information in Selection (Medicare.gov)
- Social or Functional Adjustment and Measure Type

Technical Guidance Sections

Accounting for Social and/or Functional Challenges in Provider Performance Scores (pages 4 – 5)

Focus Group Notes (Key Considerations)

- The conceptual model should consider the time period related to the social risk (i.e., how long has the patient had the social risk) and how that impacts the outcome
- It is not targeted in what they are trying to measure and the impacts of SRFs—consider the time period related to the social risk in its ability to impact that outcome.
- We need to know more on how it sits in the process and payments.
- Risk adjustment is important, even if you are being evaluated at the 30-day mark and everyone is not being evaluated at the same starting line.
- Consider people without resources. Risk adjustment may disincentivize providers for accepting patients with higher risk.
- The conceptual model should look at the whole picture. It is important to not focus on a specific snapshot.
    - Better understanding of the implication of the diversities of community is needed. ___is/are interested in the range of diversity that the clinician treats. What is the range of diversity that the clinician treats? Is the clinician a person of color as well as the patient?
    - Nonclinical data are very difficult to collect and standardize and use of proxy information.
- For LGBTQ+ elders, they may be excluded from risk models due to sexual orientation and gender identity. Performing a risk adjustment may be impossible to capture the LGBTQ+ community. For example, a lesbian elder will have a different risk than a transgender person.
- Racist policies and payment structures are different between public payment systems. Disproportionally to utilize healthcare coverage due to racism. We do not want to penalize providers who treat patients with unmet needs. There may also be challenges with patients accessing providers.
- Could there be other ways to not penalize providers? Could we flag patients with more complex social needs and then the provider can be paid a higher rate for these patients based on their risk?

Emerging Themes

- Improvements to the conceptual model
- Expanding the provider's LOC
- The future of quality measurement

Minority Viewpoint

- Risk adjustment may disincentivize providers for accepting patients with higher risk. Look at the whole picture. It is important to not focus on a specific snapshot.

Use of Quality Measure Information in Selection (Medicare.gov)

Focus Group Notes (Key Considerations)

- Stratify patients with high risk, low risk, or medium risk, and show performance in each of those areas. Identifying what is considered high, low, and medium risk is needed.
- It would be nice to be able to see the characteristics of the patients and communities that providers serve.
- There are no data on whether a provider will treat patients who are sex workers, drug users, part of the LGBTQ+ community, etc. People who are marginalized who look for access do not refer to a specific score (e.g., colonoscopy) to select a healthcare center.
- Stratification is hard to narrow down to one specific number from a rating of 1 to 5. It would be helpful to know which type of patients each provider is serving? We need to understand that there are different communities that have different health needs. If a primary care physician (PCP) is LGBTQ+ friendly, the patient may feel more comfortable with receiving care.

Emerging Themes

- Stratification

Minority Viewpoint

- None. All were in agreement that stratification is important and has a role in performance score reporting. This is aligned with the Technical Guidance.

Social or Functional Adjustment and Measure Type

Focus Group Notes (Key Considerations)

- Certain populations are less likely to utilize care due to social risk discrimination, and everyone should be accountable for these measures.
- There is a role for risk stratification, but extremely strong data collection and reporting are both needed.
- There could be reasons or justification for adjusting across all of these measures (e.g., process, outcomes, and/or cost), and we need to be careful with the data on why you are risk-adjusting.
    - Which social and functional factors have an impact that is associated with some of these measures that we are expected to see the changes in and then adjust accordingly? For specific populations with high SRFs, you may need more social workers and potentially different benchmarks and requirements based on the risk stratification.
- Process measures should not be, but outcome measures should in order to show conceptual relationships. The total cost of care is different because sometimes you want to spend more to provide better outcomes. Less is not always better, and more is what you need to provide better care.
- It would be helpful to develop a few models in parallel until we deploy out into the real world. Which model of risk adjustment is best for that population?

Emerging Themes

- Improvements to the conceptual model
- Stratification

Minority Viewpoint

- Process measures should not be, but outcome measures should in order to show conceptual relationships. The total cost of care is different because sometimes you want to spend more to provide better outcomes.

Payer and Purchaser Focus Group

Topic Areas

- Conceptualizing the Model (pages 14 – 19)
    - Intended Use (page 19)
    - Level of Measurement (i.e., Locus of Control) (pages 18 – 19)
- Considerations for Determining the Final Risk Adjustment Model (pages 25 – 28)

Technical Guidance Sections

Conceptualizing the Model (pages 14 – 19)

Focus Group Notes (Key Considerations)

- The conceptual model is critical, and measures that do not have strong conceptual models tend to fall apart; a graphic is a great way for measure developers to identify problems initially.
- Measure developer feedback is needed after the measures have been developed, submitted, and tested; maintenance as a pre-step is needed.
- Provide clarity on the conceptual model diagram, possibly a descriptor.
- Provider demographics and insurance status of patient should be considered in the conceptual model.

Emerging Themes

- Improvements to the conceptual model

Minority Viewpoint

- Accountable entity is heavily geared to providers towards care delivery as opposed to payer or the function.
    - SDOH do not have to always be delivered through a provider but can also be delivered through the health plan (e.g., Kaiser Permanente)
- The conceptual model is possibly missing chronic care perspective regarding community resources, specifically the differences in the environments (draw out more of the clinical factors and the chronic care that surrounds it as well).

Intended Use (page 19)

Focus Group Notes (Key Considerations)

- Page 19/the second paragraph is written completely agnostic; it is difficult for measure developers to move past barriers and identify where their measures may be used, and social risk adjustment may not be necessary. For the purpose of payment where hospitals want to be judged fairly because of the payment incentives versus readmission to production program where there is payment penalty. It is difficult for measure developers to measure with the correct adjustment that accounts for both scenarios.
- The Technical Guidance does not fully provide clarity for measure developers on what is being measured. For future measures, or (i.e., readmission measure), there are concerns regarding adjusting away SDOH or a need for NQF to provide guidance or potentially create a different scope of that measure or clarify the scope. We recommend considering refining the measure to create clarity in the denominator, the time and space of measure, and the components of the

measure instead of risk-adjusting. We will alleviate the mathematical versus the moral choice of trying to balance between social needs and risk adjustments.

- Some consideration for the intention of how that measure is going to be used (e.g., evaluations for hospital performance for performance improvement purposes and providing clarity on what they could do to improve care and access or internal processes for that specific set of outcomes.)
    - Consider how physicians have to apply this on a day-to-day basis. Is it provider burden or burnout? How do we operationalize it?
- Guidance on the appropriate use of adjustment versus stratification is needed. If the intended use is performance improvement, stratification could help support that.
- Perhaps there is a way to conduct both risk adjustment and stratification by looking at the overall performance of risk-adjusted ____ and then trying to look over time and at improvements on non-risk-adjusted measures.
- Make providers accountable for taking care of patients and not always risk-adjusting.
- Risk adjustment or stratification will depend on the type of measures, such as risk adjustment utilization measures at either the plan level or even the provider level. How do we get SDOH information on membership or populations? Is there any work that NQF can do on its end?
- If the measure is measuring many factors (e.g., provider activities, influences of social needs), it should have a decreased focus on risk adjustment but create a better measure. Create a measure that has a specific LOC and uses parameters. We need to be clear about the intended outcome and the benchmark of performance improvement processes. Measures that are unclear about what they are measuring should not consider social needs/social demographic information because there may be a better opportunity to be measured better.
- If we know measures have multiple loci of control, it should be measured as well. But to solely put on the provider and knowing that there are other factors, there may be an opportunity to refine the measure.
- Measure developers will include empirical evidence or understanding of how SRFs have an impact on care delivery, the outcome, and cost. Look at the empirical evidence to support decision making on whether it would be appropriate to stratify or risk-adjust.

Emerging Themes

- **Providing Clarity**
    - Create a different scope of measures or clarify the scope
    - Descriptors
- **Burden**

Minority Viewpoint

- Barriers and difficulties exist with measure developers identifying where the measures are being used.
- Risk adjustment may not be necessary for quality improvement.
- Transparency is key for whatever the intended use is, whether it is risk adjustment in certain cases or stratification in other cases.
- A bad patient-reported outcome measure (PROM) would not be a good reason for risk adjustment and would be beneficial as a tool for measure developers to use at the beginning. An

ASPE report has language specific to patient safety measures and hospital-acquired conditions; the examples are not appropriate for this type of adjustment.
- For specific clinical outcomes, specific quality measures may have different weights to different type of vulnerabilities; transportation and food insecurities may have greater numbers to address disparities.

Level of Measurement (i.e., Locus of Control) (pages 18 - 19)

Focus Group Notes (Key Considerations)
- The LOC needs to be context-specific.
- ___ is/are related to the core principles on page 9 of the guidance. For the principle related to race and ethnicity used a proxy for social risk, fix the wording, as this could perpetuate the misconception/thinking that those social needs and social risk are connected to race, and this is not the case.
- Providers have a lot of control and the role to make sure it is represented that the people they serve. Also, look at how they look at their data. We should not only decrease the readmission but also reduce disparities. Opportunity with providers and partnership and make sure equity is part of quality. How we talk about equity and start to question and do not want to share the demographics and the diversity.

Emerging Themes
- Expanding the LOC
- The future of measurement

Minority Viewpoint
- is/are related to the core principles on page 9 of the guidance. For the principle related to race and ethnicity used a proxy for social risk, fix the wording, as this could perpetuate the misconception/thinking that those social needs and social risks are connected to race, and this is not the case.

Considerations for Determining the Final Risk Adjustment Model (pages 25 – 28)

Focus Group Notes (Key Considerations)
- Shift to a benchmark model, consider the organization, and possibly use it as a model for payment incentives. There is not enough discussion in this report about bias and racism. Also, consider whether it is true that race is connected to being socially disadvantaged; is it ever appropriate for it to sit on the accountable entity? ___ was/were very hesitant to provide any and may require further thought. We want to uncover racism.
- It is uncertain whether risk adjustment is needed for these payment models.
- There is a lot that goes into the measure that impacts policy. What is appropriate to risk-adjust? What is helpful to risk-adjust that should the demographics and identify the disparities? Using risk adjustment and have not leveraged to use an any outcome. Race is not used in any risk adjustment and the social needs factors. Going back to the LOC, we felt as though there is no need to. What would it look like if we were to measure it? There is a need for data standardization when it comes to collecting real race/ethnicity data. For race, there are only six classifications, and this does not represent true patient race. Providers need support on data capture, such as an interoperability framework to create consistency across the organization.

There has not been a clear measure related to quantifying the different domains of social needs. The moment you ask about the patient, there has to be an answer to the question.

- Stratification can be done if they do not have good or accurate data.

Emerging Themes

- Expanding the LOC
- Stratification
- Risk factor selection
- Improvements to the conceptual model

Minority Viewpoint

- There is not enough discussion in this report about bias and racism. Also, consider whether it is true that race is connected to being socially disadvantaged; is it ever appropriate for it to sit on the accountable entity?