NATIONAL QUALITY FORUM
Driving measurable health improvements together

Developing and Testing Risk Adjustment Models for Social and Functional Status-Related Risk Within Healthcare Performance Measurement

FINAL TECHNICAL GUIDANCE – PHASE 2

December 21, 2022

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

Table of Contents

Executive Summary

The National Quality Forum (NQF) endorses performance measures that are intended for use in both performance improvement and accountability applications, such as public reporting and pay for performance. The overall performance measure score is used to make a conclusion about the quality of care provided by an accountable entity (i.e., a hospital, health plan, physician practice, or any other entity that is being assessed) in relation to other entities or some other comparator, such as average performance. Measures that assess performance of health outcomes are risk-adjusted to ensure the measure score is affected as little as possible by factors other than quality of care, such as differences across accountable entities in patient characteristics that are present at the start of care.

As a consensus-based entity, NQF seeks to work with quality measurement stakeholders, including patients and caregivers, clinicians and specialty societies, private and public health systems and health plans, measure developers, and technical experts, to develop standard, science-based approaches to healthcare quality measurement. However, this stakeholder community has lacked consensus on how and whether to adjust for social and functional status-related (referred to as functional risk factors in this guidance) risk factors in outcome and cost/resource use measures. Accordingly, the Centers for Medicare & Medicaid Services (CMS) contracted with NQF to develop Technical Guidance for a standard risk adjustment framework that reflects best practices for, and addresses methodological trade-offs of, adjustment for social and functional risk factors within quality measurement.

This Technical Guidance presents the result of this two-and-a-half year project. In the first year, NQF developed the Technical Guidance through an environmental scan and input from a Technical Expert Panel (TEP) and the public. NQF published the guidance in a prior version at the conclusion of the first year.[a] In the second year, NQF further vetted the Technical Guidance through a series of focus groups, ensuring broad stakeholder input that included the perspectives of the patient community, clinicians, health system administrators from both the public and private sectors, technical experts in risk adjustment methods, and NQF-convened measure evaluation Committees. A major focus of this broadened stakeholder engagement was to solicit feedback from individuals with minority viewpoints (i.e., those who disagree with the initial phase of the **Technical Guidance**). Furthermore, to support the advancement of health equity within this work, NQF recruited members of communities who 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. The feedback received was considered by the TEP, which informed revisions and enhancements to the guidance with further public input. This document presents the culmination of work done as final guidance for developing and testing risk adjustment models for social and functional risk within healthcare performance measurement.

---

[a] NQF. "*Developing and Testing Risk Adjustment Models for Social and Functional Status-Related Risk Within Healthcare Performance Measurement - Final Technical Guidance*" Aug 2021. https://www.qualityforum.org/Publications/2021/08/Developing_and_Testing_Risk_Adjustment_Models_for_Social_and_Functional_Status-Related_Risk_Within_Healthcare_Performance_Measurement_-_Final_Technical_Guidance.aspx

This work aims to further the development and use of fair, unbiased quality measures that drive healthcare quality improvements and reduce disparities in health and healthcare. There is broad agreement that measurement must support efforts to improve the equity of health outcomes and access to healthcare. Advancing equity is a widely shared priority and goal of CMS's National Quality Strategy.[1] Quality measures are a key lever for making progress toward this goal. There is also broad agreement that measures should not be biased. However, there are strong differences of opinion as to whether adjusting for social risk factors achieves these goals. Furthermore, for patients who have both social risk factors, such as poverty, and functional risk factors, such as dementia, the presence of both could increase the likelihood that these risk factors will interact, which may lead to increasing the adverse impact of risk factors on the measured outcome.[2] Therefore, functional risk factors are important to examine since they may confound or mediate the relationship between social risk and the measured health outcome.

This guidance fills a gap in measurement science by providing a standard framework for evaluating the methodological trade-offs encountered in the decision about whether to adjust for social and/or functional risk. It also provides a practical and detailed, step-by-step process for selecting and testing social and functional risk variables. As background to the key recommendations noted below, we briefly summarize here the range of views and technical considerations discussed by stakeholders that inform the Technical Guidance.

Those opposed to social risk adjustment point out two potential unintended consequences of adjusting outcome measures for a social risk factor, such as income, education, social relationships, urbanicity/rurality, or health literacy: (1) establishing a lower standard of care for patients with such risk factors and (2) obscuring differences in quality among providers. These unintended consequences could occur because some risk adjustment models estimate an expected outcome rate for each measured accountable entity after adjusting for patient risk factors (including clinical, demographic, social, and functional factors), which influence the outcome rate independent of quality. Take, for example, a measure that estimates the number of expected complications aggregated among an accountable entity's patients and compares it statistically to the entity's actual complication rate. The measure score is calculated as a ratio, such as the observed-to-expected (O/E) or predicted-to-expected (P/E) number of adverse outcomes. A lower score is better. Adjusting the measure for a social risk factor often raises the expected number of complications for patients with that factor. Increasing the denominator (E) for patients with social risk factors results in a lower O/E or P/E ratio for such patients compared with those who do not have these factors. This could be viewed as accepting a lower standard of care for patients with social risk factors.

The related effect—potentially obscuring provider quality differences—occurs because a measured accountable entity, such as a hospital in the above example, caring for more patients with social and functional risk factors can have a greater number of complications, but a similar measure score (e.g., O/E ratio) compared with an accountable entity having fewer patient complications but the other risk factors being equal. Hence, adjusting for social and functional risk factors could conceal differences in quality and lower the visibility and urgency of improving care for the most vulnerable patients, thereby slowing progress toward equity.

Conversely, those in support of social risk adjustment emphasize a potential behavior response among providers caring for populations with higher social risk: It could reduce access to care for those populations. To proponents, risk adjustment is important to preserving and expanding access to care among vulnerable populations. For example, to avoid financial penalties triggered by poor measure scores on measures not adjusted for social risk factors, measured entities might avoid caring for disadvantaged patients, who are often at inherently greater risk of worse outcomes. This potential unintended consequence of using unadjusted measures in value-based payment (VBP) programs could therefore reduce equitable access to care. Absent social risk adjustment, a provider caring for a higher proportion of socially disadvantaged patients might have a less favorable measure score if the difference in outcomes reflects social risks that quality care cannot fully address. Additionally, supporters of adjusting for social risk factors also posit that such adjustments are necessary to ensure fair comparisons of performance across accountable entities.

In addition, there is disagreement on how the types of accountable entities measured and the intended use of the measure in quality reporting and payment programs should shape decisions on whether to adjust for social and/or functional risk factors. Stakeholders agreed that different types of accountable entities—from individual providers to large health plans—have variable ability to address patients' social needs. However, the TEP and other stakeholders, including focus group participants, expressed a range of viewpoints on the actions accountable entities should be taking, given the resources they have or may be provided through incentive programs to help mitigate the adverse influences of their patient's social risk factors on health outcomes.

There is also disagreement on the extent to which quality measures should be designed to incentivize and reflect the provision of more holistic patient care that better addresses the social determinants of health (SDOH). In recent years, healthcare providers have demonstrated they can reduce the health impacts of risk factors, such as poverty, transportation, and nutrition, by assessing and mitigating the ways patients' social risk impedes good outcomes.[3,4] This is an area of particular focus for providers in systems that have accepted financial risk for the cost and/or outcomes of care, such as accountable care organizations (ACOs) and bundled payment models, because they recognize that addressing these social risk factors contributes to their success at managing the total cost of care and improving patients' health outcomes.[5,6] However, addressing the risk of poor outcomes conferred by social risk factors may require additional resources and collaboration with community-based organizations, which may not always be feasible. Therefore, addressing the risk of poor outcomes due to social risk factors is not uniformly viewed as an integral aspect of providing quality clinical care, and it may not be practical for all providers. Measurers can take a broader view but should acknowledge that not everything is within the locus of control of the accountable entity. However, the use of risk adjustment methodologies in performance measurement alone should not be used to achieve resource (re)allocations.[7]

Similarly, stakeholders acknowledge that determining whether measures should be adjusted for social risk factors will vary depending on each specific measure and the structure of the VBP or quality reporting program in which a measure is used. Measure specifications, including risk adjustment, and the design of the program in which the measure is used can inform each other. However, if a risk-adjusted measure is used by multiple users, the program design (e.g., VBP models) may not be identical

across the measure users. Therefore, in some cases, programs may account for social risk by adjusting payments or using peer grouping rather than using a risk adjustment methodology in the measure specifications. Advocates for adjusting payments or deploying peer grouping, which is more commonly used in the public sector, believe that these approaches can spur further investments in health equity and mitigate providers' incentives to avoid populations with high social risk. For example, the Medicare Hospital Readmissions Reduction Program (HRRP) uses peer grouping to facilitate comparisons of hospitals based on the proportion of dual Medicare-Medicaid beneficiaries that they care for.[8] In this approach, performance scores for accountable entities are neither adjusted nor stratified. Thus, penalties are applied on the basis of performance within peer groups rather than on the basis of performance relative to the entire universe of accountable entities, which mitigates the need for upstream risk adjustment of the measures.[9] Another example is the Medicare ACO Realizing Equity, Access, and Community Health (ACO REACH) Model, which is a redesign of the Global and Professional Direct Contracting Model. The ACO REACH is designed to promote health equity and address healthcare disparities by incorporating new health equity-focused elements affecting payments, data collection, provider selection, and care delivery.[10] New to the ACO REACH Model, CMS will apply a health equity adjustment for benchmarking for all ACO types starting in 2023. These adjustments will be made at the beneficiary level using dual-eligibility status and the Area Deprivation Index.[11]

In the private sector, payers commonly leverage their VBP models to promote health equity by incentivizing the reduction in health disparities and mitigating the impacts of social risk factors.[12] This can range from rewarding data collection and referrals to community-based organizations to working with providers to offer their patients transportation services. However, while these private programs commonly offer resources, they are not all structured to account for risk at the program level, and thus, they may benefit from leveraging measures that are adjusted. Moreover, some of the programs that do make payment adjustments may find that such an approach may not sufficiently adjust for all factors outside of the providers' control. Thus, measure developers should consider, if known, the program context in which the measure is expected to be used.

This project sought to develop a cohesive framework for considering social and functional risk adjustment, informed by technical, empirical, and policy considerations, and with input from an environmental scan, a TEP, stakeholder focus groups, and the public. Based on these inputs and the iterative comments from the TEP and stakeholders, NQF developed this final Technical Guidance that facilitates structured consideration of social and functional risk adjustment within quality measurement; provides best practice guidelines; and, when possible, gives more specific direction on emerging issues. Where appropriate, it provides flexibility to accommodate diverse measures and uses. In addition, this document acknowledges divergent points of view and seeks to minimize the burden on measure developers. Specifically, the Technical Guidance:

- articulates core principles for risk adjustment;
- sets forth a five-step process to developing and testing risk models for outcome and cost/resource use measures for NQF endorsement review, beginning with building a conceptual model that illustrates the potential pathways between the social and/or functional risk factors, patient clinical risk factors, quality of care, and the measured health outcome;

- defines seven risk adjustment, best practice standards and integrates them into the five-step process;
- describes a set of social and functional risk factors that developers should consider, at a minimum, within the overall risk adjustment strategy;
- discusses key considerations for identifying and selecting risk factors within data sources of sufficient quality, including the strengths and limitations of common and emerging data sources (e.g., electronic health records [EHRs], claims);
- articulates guidance for stratification analyses, including a minimum set of stratification variables; and
- provides a path forward for the field that will advance equity while ensuring a transparent, consistent, and fair approach to measurement.

Based on the input from project stakeholders, this guidance advances consensus and further delineates best practices for social and functional risk adjustment within quality measurement in several key areas:

1. This guidance provides a standard framework that supports consistent development and review of risk adjustment models and the use of social and functional risk variables in these models. It encourages developers to choose the most appropriate risk variables to be included in the risk model based on the specific measure and program context. Specifically, to inform risk variable selection, **developers should prepare a conceptual model that illustrates the potential pathways between the social and/or functional risk factors, patient clinical factors, healthcare processes (e.g., care delivery steps, whether certain tests were conducted, care coordination), and the measured outcome, recognizing that often, these are not completely understood. The rationale for risk adjustment variable selection should derive from the relationships illustrated by the conceptual model**, which shows potential mechanisms by which the risk factors may affect the measured outcome. The conceptual model should explicitly discuss assumptions made about factors that influence the outcome, including which factors substantially reflect potential actions by the entity (i.e., the accountable entity can take action to influence and change the factor's effect on the outcome). More specifically, the conceptual model should delineate what intrinsic patient factors might influence the health of the patient when they initially present for care (e.g., language), and which are therefore appropriate for adjustment, versus provider factors related to the quality and effectiveness of care (e.g., providing language or translation services), which are not appropriate for adjustment. Assumptions about the accountable entity's ability to influence social and/or functional risk factors should be articulated. **Developers should also examine the role of social and/or functional risk factors in the context of the measure's expected use, if known, as the measure's intended use may affect decision making regarding risk variable inclusion, feasibility of stratification, and potential unintended consequences.** Statistical significance testing of social or functional risk factor variables is a critically important consideration, although it should not solely determine the inclusion or exclusion of a risk factor within the final risk adjustment model.

2. This guidance establishes **a minimum set of social and functional risk factors that must be considered within the conceptual model. The guidance recognizes that variable selection for**

risk model development should be informed by both the conceptual model and the availability of data of sufficient quality. There may be other social and/or functional risk factors identified beyond the minimum set defined within this guidance.

3. **Race is qualitatively different from other social risk factors because the race variable often reflects a broad range of influences**, including socioeconomic/sociodemographic status, environmental factors, access to high quality care, bias and discrimination, genetic and epigenetic predispositions to certain diseases, and different responses to treatment (e.g., pharmacogenomics). The contributions of these various components to the race variable differ among specific disease processes and risk models, and the precise proportion of each is often unknown. Further contributing to the knowledge gap in this area, patients with non-White race have often been underrepresented in biomedical research and genome-wide association studies. Race also carries information about other social risk factors that correlate with race, such as income; yet there is still limited availability of robust socioeconomic status (SES) indicators for which race may sometimes be a proxy. The persistent association of Black race with worse health outcomes in risk-adjusted models likely reflects, among other factors, the collective health effects of profound historic and continuing discrimination, including bias within the healthcare system.[13]

   Given all these considerations, as well as the ongoing societal and scientific discussions on these topics, it is not surprising that consensus regarding the role of race in risk models was not achieved by the TEP. Some members of the TEP and focus groups oppose adjusting for race in risk models due to concerns that this could effectively set lower standards for minority populations and inadvertently perpetuate long-standing disparities for Black patients and other at-risk racial minorities. However, other TEP members emphasized the importance of adjusting for race due to its importance as a risk predictor. Furthermore, adjusting for race may also be important to ensure that accountable entities who care for a high proportion of at-risk patients are not penalized by measures when they are used in VBP programs that do not provide the additional resources that may be required for achieving good quality outcomes for socially vulnerable patient populations.

   Hence, this guidance does not state whether adjustment for race is or is not appropriate, as the TEP preferred to articulate both sides of the issue and allow for the **consideration of adjustment for race on a case-by-case basis, given the conceptual model and the features of the VBP program in which the measures may be used, if known.**

4. Lastly, stratification is a critical tool for reporting on differences in measure results between subgroups of patients and thus identifying potential disparities in care. Stratification can reveal differences in outcomes across subgroups with different patient characteristics known to be associated with disparities. This guidance defines several key variables for stratification analyses, which were intentionally limited to minimize developer burden. **The TEP agreed that race, ethnicity, an indicator of urbanicity/rurality, an indicator of poverty, and an indicator of**

disability status were the variables that developers should analyze for stratification, at a minimum.

Risk adjustment is not perfect, and the same limitations of clinical risk adjustment also apply to adjustment for social and functional risk factors (i.e., risk adjustment can only account for measurable and available factors). The intent of this guidance is to advance consensus on best practices for risk adjustment models, which will promote the development and endorsement of fair measures that help to advance healthcare equity. This guidance will facilitate a consistent and more widely accepted approach to risk adjustment for social, functional, and other risk factors.

Introduction

Background

Over the last decade, the quality measurement enterprise has rapidly moved toward linking payment to quality of care, generally known as VBP, to improve healthcare delivery and health system accountability. This improvement is realized through quality, efficiency, safety, and patient experience. For VBP to be successful, accountable entities need comprehensive, reliable, accurate, and timely information to make quality care decisions that result in improved outcomes for patients while being held accountable for those outcomes in a fair and unbiased manner. To account for differences in the patient population across accountable entities, **risk adjustment** methods have been applied to many quality performance measures, but not all, and not in a standardized manner across measures.[14] Furthermore, patients need accurate and reliable information on the performance of **accountable entities** (e.g., clinicians, health plans, and health systems/hospitals) to make informed care decisions.

Risk-adjusting outcome and cost/resource use measures to account for differences in patient health status and clinical factors (e.g., comorbidities, severity of illness) that are present at the start of care has been widely accepted and implemented.[15,16] However, the increased use of outcome and cost/resource use measures in payment models and public reporting programs has raised concerns regarding the adequacy and fairness of the risk adjustment methodologies used in these measures, especially as it relates to **functional status**-related risk factors (referred to hereafter as *functional risk factors*), such as the ability to perform activities of daily living (ADLs) (e.g., eating, bathing, dressing, and toileting[17–19]) and **social risk factors**, such as income, education, social support, neighborhood deprivation, and urbanicity/rurality.[20,21] Functional risk factors are important to examine because the presence of both social and functional risk could increase the likelihood that these risk factors will interact, which may lead to increasing the adverse impact of risk factors on the measured outcome.[2]

The relationships between social, economic, and environmental risk factors and health and health-related outcomes as well as the unequal burden of these risks across sociodemographic groups (e.g., race, ethnicity, language preference, disability status, sexuality and gender identity, and rural subgroups) have become even more apparent during the coronavirus disease 2019 (COVID-19) pandemic.[22–24] The root causes of inequities in exposure, access to testing, and treatment and outcomes are multiple and often interrelated. The impact of social and functional risk factors on health and healthcare outcomes highlights the importance of recognizing and appropriately considering all

applicable clinical, social, and functional risk factors when reporting and evaluating quality measures and accountable-entity performance. These risk factors should be informed by a conceptual model for the measure under development, as not all clinical, social, and functional risk factors considered would be appropriate for inclusion in a final risk model for a performance measure. The COVID-19 pandemic underscores the importance of exploring and appropriately adjusting for applicable risk factors (i.e., clinical, social, and functional) to ensure accurate assessment and prevent inappropriate financial penalization of accountable entities due to caring for patient populations with increased social and/or functional risk.[25]

Health equity is fundamental to all quality improvement efforts. Quality measurement should contribute to closing the health equity gap and not inadvertently institutionalize or exacerbate it. To that end, CMS has prioritized the advancement of health equity within its National Quality Strategy and is focused on leveraging quality measures to promote health equity and reduce disparities in care. In alignment with CMS' health equity goals, NQF applies an equity lens to every aspect of its work, with the goal of empowering healthcare stakeholders to take meaningful and measurable action to achieve health equity.[26] This includes addressing quality and measurement gaps in key national health priorities, including the endorsement of performance measures that can identify and have the potential to reduce **health disparities**. Addressing the wide spectrum of disparities must be considered as a key component for successful health outcomes across the nation. As social and functional risks are increasingly realized as having an impact on health and healthcare outcomes, NQF recognizes that fully addressing inequities associated with race, ethnicity, or social risks requires a holistic policy approach and a private-public sector partnership that goes well beyond the purview of quality measurement.

This Technical Guidance acknowledges the holistic approach needed to address health equity and focuses on a specific measurement science debate: whether and how to adjust healthcare performance measures for social and functional risk factors so that accountable entities will be compared fairly. There is a clear distinction between directly adjusting payment rates for social risk factors and adjusting quality measure scores that may be tied to financial bonuses and incentives. When there is a conceptual rationale for social risk factor adjustment, there is an option to directly adjust the payment rate or to adjust the quality performance score. Either approach may be appropriate depending on the goals and context of the program. This guidance focuses on the latter case, that being risk-adjusting quality measures. However, quality measure adjustment alone cannot and should not be used to achieve resource (re)allocations. For example, social and/or functional risk adjustment of patient-level factors does not address potential differences in community factors, such as public funding or area healthcare resources, which may have a substantial impact on accountable entity performance. Similarly, risk adjustment of healthcare performance measures alone cannot be the only approach to addressing the health system goal of health equity.

Yet, as discussed in the *Executive Summary*, there are different points of view about whether adjusting for social risk factors that influence measured outcomes will accelerate or hinder efforts to improve health equity. Those in support of social risk factor adjustment posit that by not adjusting for social risk within certain performance measures, accountable entities may avoid caring for the most at-risk and disadvantaged patients because of their anticipated worse outcomes or higher costs, potentially

worsening inequities. Alternatively, those opposed to the inclusion of social risk factors in risk adjustment models argue that adjusting certain performance measures for social risk may not make transparent the differences in care outcomes. Because of the complexity of these issues and the associated robust national debate, white papers and guidance documents have been published by various organizations, including NQF; the National Academies of Sciences, Engineering, and Medicine (NASEM); and the Assistant Secretary for Planning and Evaluation (ASPE).[7,27,28]

Prior to 2014, NQF's measure evaluation guidance prohibited the inclusion of social risk factors in the risk adjustment models of measures submitted for NQF review and endorsement due to concerns of masking inequities in care.[7] In 2014, NQF convened a Risk Adjustment Expert Panel, which recommended allowing risk adjustment when a conceptual rationale and empirical relationship is present.[16] The NQF Board of Directors implemented a trial period in 2015, during which adjusting measures for social risk factors was no longer prohibited.[29] At the conclusion of the trial period in 2017, NQF Standing Committees and measure developers reiterated the importance of addressing all factors (both clinical and social) that can influence the result and the validity of a performance measure in truly reflecting care quality and resource use.[30] These efforts have demonstrated that there is still debate on whether and how social risk adjustment should be considered for candidate measures for NQF endorsement. Measure developers are still challenged with obtaining granular data that accurately reflect a person's social risk. Additionally, functional risk factors have been underutilized. Nevertheless, they play a critical role in risk adjustment since they may mediate the relationship between social risk, quality outcomes, and resource use.

Measure developers, stewards, and program implementers have long expressed a need for technical guidance and standardization in developing, testing, and evaluating risk adjustment models for measured outcomes affected by social and/or functional risk. Approaches to risk-adjusting these outcome measures require consideration of the data sources and statistical models used; the specific risk factors used to represent functional status, SDOH, SES, and sociodemographic status (SDS); and how to determine whether these factors should be included in the overall risk model. These considerations need to take into account the potential increased burden on measure developers in terms of increased testing requirements.

Hence, developing a standardized, consistent approach to risk adjustment would facilitate more consistent assessments of the role of functional, social, and clinical risks; enable fair, unbiased comparisons of the performance of the accountable entities with different patient case mix; and report and monitor disparities across subgroups.[30]

Purpose

This Technical Guidance provides quality measure developers with a standard risk adjustment framework, articulating a step-by-step approach for developing risk adjustment models that consider social and/or functional risk factors for outcome and cost/resource use performance measures. This guidance considers the strengths and limitations of developing these risk models, including the commonly used methods and practices, the availability of data sources, and potential policy implications. Through input from an NQF-convened TEP, this document identifies standard best

practices that developers should do, at a minimum, for social and/or functional risk adjustment within performance measure development.

The intent of this Technical Guidance is to serve as a resource for both novice and experienced measure developers. It will further facilitate consistency in the evaluation of risk adjustment models within performance measures for NQF endorsement. Therefore, this guidance ***does not*** describe recommendations for risk-adjusting these factors beyond the scope of NQF endorsement, namely, whether outcome and resource use measures should be adjusted for social and functional risk at the program level. For example, the **21st Century Cures Act** required the HRRP to use a stratified, or peer-grouping, methodology to evaluate hospital performance relative to other hospitals with similar proportions of patients who are dually eligible for Medicare and full-benefit Medicaid.

Although it is uncommon, there may also be a relationship between social/functional risk factors and some processes of care that suggests the need for risk-adjusted process measures (e.g., filling a drug prescription could be affected by a patient's SES as in **NQF #0541**, which adjusts for age, gender, low-income subsidy [LIS]/dual status, and disability status). This guidance, although not directed at process measures, can inform that assessment. Certain measures, such as serious reportable events (SREs) or never events, should not be risk-adjusted for social and/or functional risk factors since they are largely preventable and indicative of a problem in a healthcare setting's safety systems. Instead, these measures should be stratified by social and functional risk factors to facilitate the reporting of these serious reportable events in subgroups.

Project Overview

With a goal of advancing consensus on and methods for measurement science, NQF developed this Technical Guidance for measure developers; it includes standard best practices as requirements developers should follow, at a minimum (referred to hereafter as *minimum standards*), for social and/or functional risk factor adjustment in quality performance measure development. To inform this work, NQF, with support from CMS, convened a **multistakeholder TEP** beginning in the fall of 2020 to provide input and guidance on the current state of risk adjustment for social and functional status in measurement, comprehensive recommendations and methods that experts agree are best practices for social and functional risk adjustment, the appropriateness of a standard risk adjustment framework, and the development of step-by-step technical guidance for measure developers.

NQF conducted this project in two phases, both supported by CMS. During the first phase of this effort (2020–2021), the TEP provided guidance on an NQF-conducted **environmental scan**. The scan identified and assessed the current state of data sets used for the risk adjustment of functional and/or social risk within quality measurement, the conceptual and statistical methods used, and the approaches to interpretation and the decisions to include or not include functional and/or social risk factors within the final risk adjustment model. Additionally, the environmental scan considered the scientific acceptability of any standardized risk adjustment framework. NQF used the results of the environmental scan, together with the input and diverse perspectives shared by the TEP, to develop the initial phase of the **Technical Guidance**. This guidance describes the process of conceptualizing an outcome or a cost/resource use performance measure, the subsequent risk adjustment model development

(specifically accounting for social and/or functional risk), and the decision making that will be needed for NQF endorsement review.

During the second phase of the project (2021–2022), NQF reconvened the TEP (**Appendix A**), which largely consisted of members from the initial phase of this work and some new members, to update the Technical Guidance based on broader stakeholder perspectives. To accomplish this goal, NQF socialized the guidance with various healthcare quality stakeholders through two CMS-convened meetings and six NQF-convened focus groups. The two CMS-convened meetings were the Measure and Instrument Development and Support (MIDS) C3 Forum and the Quality Measurement Technical Forum (QMTF). During these two meetings, NQF presented findings from the initial phase of this work, including the environmental scan and elements of the Technical Guidance, namely the consensus standards and the conceptual model. For the focus groups, NQF recruited individuals representing healthcare quality 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., the Scientific Methods Panel [SMP], Standing Committees, and the Consensus Standards Approval Committee [CSAC]).

In particular, NQF reached out to individuals with minority viewpoints (i.e., those who disagree with the Technical Guidance recommendations and/or minimum standards) to elicit their input and rationale for the TEP's consideration. Furthermore, consistent with the **White House Executive Order** to advance racial equity and to support underserved communities through the federal government, NQF recruited individuals of communities that have been historically underserved. These communities included, but were not limited to, racial and ethnic minorities, individuals with disabilities, those who live in rural areas, individuals from the LGBTQI+ community, and those otherwise adversely affected by persistent poverty or inequality.

The intent of these focus groups was to gain input on how the Technical Guidance should consider the intended use of quality measures and payment programs. NQF further sought to gain feedback on the Technical Guidance's minimum standards and the conceptual approach that formed the basis for the standards. For example, the groups discussed whether the approach to risk adjustment for individual measures should be examined in the context of a measure's intended use, how to consider the accountable entity's ability to influence the impact of social and functional risk factors on the measured outcome(s) (referred to as **locus of control**), and how risk adjustment might vary by the types of outcomes being measured. Lastly, NQF sought to garner input on whether and when adjustment should occur at the measure or program level. Key considerations from this stakeholder engagement were encapsulated in the Stakeholder Feedback Memo (**Appendix F**) and were shared with the TEP to facilitate the updates to this Technical Guidance.

Key Terms and Definitions

Accountable entity refers to an individual health professional, health facility, health plan, or health organization/facility that is responsible or accountable for healthcare quality, outcomes, or cost of 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 quality among populations or social groupings (race, ethnicity, language, etc.).[31]

Functional status has various definitions in the health field. Generally, functional status refers to an attribute that assesses how a health condition has had an impact on an individual's body function, body structures, and ability to participate in activities and complete basic daily tasks.[32] Functional status encompasses both the individual's ability to carry out ADLs and to participate in life situations and society.[33] This includes basic physical and cognitive activities, such as walking or reaching, focusing attention, and communicating, as well as routine ADLs, including eating, bathing, dressing, transferring, and toileting. This also includes life situations, such as school or play for children, and for adults, working outside the home or maintaining a household. Furthermore, functional limitations occur when a person's capacity to carry out such activities or performance of such activities is compromised due to a health condition or injury and is not compensated by environmental factors (including physical, social, and attitudinal mediators). Functional status encompasses the whole person and is affected by physical, developmental, behavioral, emotional, social, and environmental conditions.[32]

Functional status-related risk adjustment refers to statistical adjustment for functional status-related variables (e.g., frailty, disability, ADLs, and cognitive function).

Healthcare disparities refer to the differences between groups in health insurance coverage, access to and use of care, and quality of healthcare services.[34]

Health disparities refer to a higher burden of illness, injury, disability, or mortality experienced by one group relative to another.[34]

Health equity is the principle underlying a commitment to reduce—and ultimately eliminate—disparities in health and healthcare and in their determinants, including social determinants. Health equity strives to ensure everyone has a fair and just opportunity to be as healthy as possible. This requires removing obstacles to health, such as poverty, discrimination, and their consequences, including powerlessness and lack of access to good jobs with fair pay, quality education and housing, safe environments, and healthcare.[35]

Locus of control refers to the scope of actions that the accountable entity can take to influence the measured outcome.[36] For example, clinicians caring for patients in small rural practices may have a limited ability to address and reduce the impact of social risk factors, such as transportation barriers and access to healthy foods, whereas larger, more urban providers, such as ACOs, may be both incentivized and resourced to address these needs.

Quality of care refers to a measure of performance on the six Institute of Medicine (IOM)-specified healthcare aims: (1) safety, (2) timeliness, (3) effectiveness, (4) efficiency, (5) equity, and (6) patient-centeredness.[18]

Risk adjustment (also known as case-mix adjustment), in the context of quality measurement, refers to statistical methods to control or account for patient- and/or community-level factors when computing performance measure scores; methods include modeling techniques, indirect standardization, or direct standardization. These methods can be used to produce a ratio of O/E, a risk-adjusted rate, or another estimate of performance. Methods include, but are not limited to, adjustment for mean within reporting unit differences in multivariable models with reporting unit fixed effects, indirect standardization, direct standardization, and matched cohort comparisons.[14]

Social drivers of health (SDOH) (also known as social determinants of health) are the social, nonmedical conditions that determine healthcare provision and health outcomes.[28] They can either improve or worsen an individual's health. SDOH and social risk factors are connected: SDOH can impact a person's health for better or worse, depending on the social circumstances. However, when those social circumstances are adverse, some people may be at greater risk for poor health. These circumstances are referred to as social risk factors.

Social risk adjustment refers to statistical adjustment for social variables, including those that are socioeconomic (e.g., income, education, occupation) and sociodemographic (e.g., age, race, gender, gender identity, Medicare/Medicaid dual eligibility, language).

Social risk factors are broadly defined to include the social conditions or factors that have a conceptual and empirical relationship to healthcare outcomes.[28] Illustratively, these factors may include socioeconomic position or status (e.g., income, education, and occupation), other cultural context, social relationships, residential and community environments, urbanicity/rurality, and health literacy. Consistent with the NASEM and ASPE reports,[27] this guidance considers a variety of sociodemographic factors as social risk factors, including age, race, gender, gender identity, Medicare/Medicaid dual eligibility, language, and uninsured status. Throughout this guidance, age is treated as both a clinical and social risk factor.[28]

Stratification refers to an approach to identifying disparities. In addition to reporting overall performance, stratification consists of computing performance separately for different strata or groupings of patients based on some patient-level characteristic(s), namely those that are social and functional status-related for the purposes of this guidance. Thus, each accountable entity has multiple performance scores, one for each stratum rather than one overall performance score.[7]

Core Principles

To ground this Technical Guidance on social and functional risk adjustment, the TEP agreed on a set of core principles. These core principles have been developed from previous NQF Technical Guidance related to two NQF reports titled *Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors* and *A Roadmap for Promoting Health Equity and Eliminating Disparities: The Four I's for Health Equity*. The principles, although grounded in sound measurement science methods, are not intended to imply a particular direction for recommendations related to risk adjustment for social and/or functional status risk. Rather, they represent a baseline of agreement on the key issues that must be considered in making recommendations. The core principles are as follows:

Core Principles:

- Performance measurement is critical to advancing quality, as articulated in the **CMS Quality Measurement Action Plan**.[37]
- Disparities in health and healthcare should be identified and reduced.
- Performance measurement should not lead to increased disparities in health and healthcare.
- Outcomes (including cost/resource use) may be influenced by patient health status and clinical, functional, and social factors, in addition to the quality and effectiveness of healthcare services, treatments, and interventions.
- Performance measures that are influenced by factors other than the care received, particularly outcomes and cost/resource, need to be adjusted and stratified for relevant differences in patient case mix to avoid incorrect inferences about performance.
- Performance measurement and risk adjustment must be based on sound measurement science.
- Risk adjustment may be constrained by data limitations and/or data collection burden.
- The methods, factors, and rationale for risk adjustment should be transparent. Additionally, the statistical approaches identified within this guidance are not intended to be overly prescriptive, as to limit the use of novel methods or to add significant burden to measure developers.
- Race as a sociodemographic risk factor may reflect complex, and at times multiple, underlying concepts. Race is correlated with other elements of social risks, such as environment, access to high quality care, and genetically mediated predispositions to certain diseases and/or different responses to treatment (including medications). Thus, the examination of a race variable requires distinguishing among the potential mechanisms through which this variable may be associated with health risk, including: (1) evidence-based genetic differences in the risk of a clinical outcome; (2) discrimination based on race (i.e., racism); and (3) as a proxy for other social structures, affects, and residual differences. Use of a race variable in risk adjustment models requires careful scrutiny regarding these mechanisms.

Environmental Scan Findings

Performance measures have been used to drive quality improvement and will continue to relate payment to the quality of care provided. The **environmental scan** revealed that common data sources used to calculate the measure and for social and/or functional status risk factor analyses include the American Community Survey (ACS), Medicare Enrollment Database, and Medicare administrative claims.[38] Commonly used methods include assessing the variation in prevalence of the risk factor across measured entities, empirically testing the association between the factor and the outcome, testing the incremental effect of risk factors in a multivariable model, assessing the adequacy of the risk model, and examining the correlation of the social/functional status risk score with the measure scores. Additionally, assessments of the contribution of social and/or functional risk factors to the risk model fit and the correlation of social or functional status-adjusted risk score and comparable unadjusted scores were both common approaches for determining the inclusion of social and/or functional risk factors

within the final risk model. Yet very different methods have been used across similar measures without consideration to a conceptual model or to a framework, which highlights the need to mitigate the existing variability and the lack of clear guidance for social and functional risk adjustment. Therefore, this TEP-informed Technical Guidance addresses this need by highlighting agreed-upon best practices as minimum standards that should be considered for social and/or functional risk adjustment within outcome and cost/resource use measurement.

Technical Guidance

Overview

This Technical Guidance serves as a resource for risk adjustment model development and testing, which accounts for social and/or functional risks. It will help guide measure developers to conceptualize, create, test, and consider risk adjustment models for performance measurement. To support developers' decision making, the guidance articulates potential trade-offs that developers should consider. Furthermore, this guidance evolves prior NQF guidance [7] and complements existing recommendations from various organizations, including ASPE and NASEM, by instructing developers to consider the risk adjustment of social and functional risks through five structured steps that implement a standard risk adjustment framework:

1. Conceptualizing the Model
2. Identifying and Selecting Potential Data Sources and Variables
3. Empirically Testing Risk Factors
4. Empirically Testing the Adequacy of the Risk Model
5. Considerations for Determining the Final Risk Adjustment Model

The Technical Guidance integrates into each of these steps a series of standards that developers should perform, at a minimum. These standards incorporate recommendations and analyses that are best practices agreed upon by experts as preferred methods for risk adjustment models.

Technical Guidance Steps

As a first step, the guidance recommends that developers build a robust conceptual model that defines the relationship between all risk variables and a given outcome, including social and functional variables. A conceptual model is important because measure developers can clearly communicate the evidence base and assumptions by mapping relationships between the risk variables and a given outcome and the mediators of those relationships. This conceptual model is critical to guide and justify developers' decisions about the final risk adjustment model and to inform empiric testing of any risk factors to help inform their inclusion in or exclusion from a model. Starting with the conceptual model, developers should consider and present their assumptions about the full picture, namely how the patient-level clinical, functional, and social risk factors, which are present at the start of care (i.e., measurement period), influence the measured outcome and how, if at all, the accountable entity can mitigate these factors to lower risk. More specifically, the conceptual model should delineate which factors can influence the health of the patient on their presentation for care versus the factors that might need to be considered in order to deliver truly patient-centered and effective care. The developer must examine

the role of social and/or functional risk factors in the context of the specific intended use of the measure and the provider's locus of control.

In the second step, the developer identifies and selects data sources and variables for inclusion in the model. By disclosing the characteristics of selected data sources (e.g., type of data [e.g., claims, registry, survey, EHR], dates of data collection), the developer facilitates transparency and helps reviewers understand the various data quality considerations, including the potential bias that may be introduced due to data availability challenges. A clear explanation of the developer's final choices and the rationale for using the selected data sources or samples is essential. It is also important to understand and consider the reliability, validity, completeness, comprehensiveness, timeliness, and generalizability of the data elements considered for inclusion in the model (see **Table 1**).

In the third and fourth steps, the developer empirically tests risk variables and the overall adequacy of the model, respectively. Empirically testing individual risk variables for their strength of association with the outcome and their distribution across measured entities informs model variable selection. Testing for validity and reliability of proxy variables and explaining how they capture the intent of the risk factor in the conceptual model is also important for transparency and understanding. The Technical Guidance reviews testing methodologies for statistically analyzing risk factors for inclusion in the model and for testing the overall adequacy of the model (i.e., calibration and discrimination tests of the risk adjustment model in subgroups specific to the measure). The guidance notes that simple bivariate and multivariable tests alone ***should not*** determine whether a social or functional risk factor is included in the risk model. Rather, there are several empirical testing methods that may be used to test the relationships described within the conceptual model. Some of these methods have been added as illustrative examples from NQF-endorsed measures (**Appendix D**). Testing the relationships in the conceptual model offers evidence of whether hypothesized pathways are appearing as significant in the model, whether pertinent subpopulations are receiving the intended attention from the design of the model, and whether there are any significant issues with assumptions made in prior steps about the model. Empirically testing model performance with and without social and functional risk factors, and also in relevant subpopulations, provides evidence of a robust and sufficiently valid risk adjustment model. While no model will be perfect, the totality of these model performance studies will provide important information on model validity to multistakeholder measure evaluation Committees.

In alignment with the core principle to avoid being overly prescriptive, the testing approaches discussed in this document are presented as guidance and are not intended to limit other testing strategies. This guidance is not intended to inhibit the use of novel methods or to add significant burden to measure developers. However, this guidance establishes a new minimum testing standard that requires measures that are risk-adjusted for social and/or functional risk to also be stratified by the risk factor(s) of interest rather than solely being adjusted. By applying stratification, the measure can be calculated and reported for patients with and without the risk factor separately in order to reveal health disparities and differential outcomes in the measure score. This requirement aligns with national efforts to leverage quality measurement to promote health equity[37] and to further mitigate concerns that risk adjustment of social and functional risks will mask disparities.

In the fifth and final step, the developer reviews the conceptual model and empirical results and considers the full range of implications in deciding whether to adjust for social and/or functional risk factors. It is important for the developer to define and articulate their rationale for the final model so others (i.e., measure evaluation Committees) may understand the decision-making process. This guidance underscores that empirical testing alone is not deterministic of whether to risk-adjust the measure and what variables to include in the final model. Therefore, developers should justify their final approach based on both empirical tests and non-empiric inputs, such as the narrative that accompanies the conceptual model. Developers should also note any potential unintended consequences of their final risk adjustment model. Failure to address risk adjustment in a systematic manner can lead to biased conclusions that may adversely affect decision making in research and health policy contexts. In summary, developers should take a balanced and thorough approach to the consideration of the trade-offs in adjusting for social and/or functional risks.

In summary, the Technical Guidance advances two new requirements for submitted measures and provides further guidance on approaches to the testing and inclusion of social and functional risk factors in risk adjustment models. The guidance requires the risk adjustment model to be grounded in a well-defined conceptual model; it also requires developers to specify a stratification approach of any measure that is risk-adjusted for social and/or functional risk factors. The guidance further informs but does not prescribe empirical approaches to testing. Rather, when a risk factor has been identified within the conceptual model, using statistical significance testing for social or functional risk factor variables should not be the sole determinant for including or excluding that factor within the final risk adjustment model. Lastly, the guidance frames and clarifies in greater detail the additional issues that developers should address in deciding when and whether to adjust for social and functional risk factors to ensure measures advance equitable care.

As the field of quality measurement changes rapidly, this guidance will also need to evolve to account for advancements in measurement science. Taking this into account, this guidance acknowledges several emerging data sources and data standardization efforts that will enable advances in quality measurement, including but not limited to Z codes within administrative claims data, EHRs, and the work conducted by the Gravity Project and the Post-Acute Care Interoperability (PACIO) Project (see the section titled **Common and Emerging Data Sources** for more information). Because risk adjustment methodology and guidance are dependent on data capture for the adjustment of social and/or functional risk, these emerging data sources will have an impact on risk adjustment capabilities in the future.

This Technical Guidance outlines best practices and a systematic process for developing a risk adjustment model with consideration for social and functional risk factors for use in outcome and resource use measures that could meet NQF's measure endorsement criteria. This guidance assists developers with providing the type of information that NQF multistakeholder Committees need to evaluate each unique measure fairly while applying a consistent standard across measures that vary widely by measure type (e.g., outcome measures, patient-reported outcome measures, and resource use measures), intended use, the outcome measured, and clinical area. However, this Technical Guidance is not intended to be overly prescriptive to limit novel approaches or innovation in the field.

Standard Risk Adjustment Framework (i.e., Minimum Standards)

This guidance identifies comprehensive recommendations and analyses that are best practices agreed upon by experts as preferred methods for risk adjustment models. The guidance refers to these as minimum standards, supporting each of the steps in this process. These standards form a framework for the risk adjustment of health outcomes and offer a robust path forward to achieve reliable and valid measure scores that can be compared across accountable entities while also considering the constraints 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. This guidance highlights the minimum acceptable standards necessary for developing meaningful and accurate risk adjustment models that account for social and/or functional risk. Additionally, this guide includes several examples of approaches and methods that help to illustrate the various steps and minimum standards in the risk adjustment process. These examples have been pulled from performance measures that have been evaluated by NQF's Consensus Development Process (CDP) (**Appendix D**) and were identified during the environmental scan measure review.

NQF considered the burden for measure developers related to the requirements for social and/or functional risk adjustment. The standards attempt to balance measurement theory with the practical constraints of measure development. Specifically, barriers to measure development may include limited data availability of the necessary risk factor variables, limited research regarding the impact of a risk factor on an outcome, or budgetary implications.

NQF also considered the burden on measure users, given risk adjustment approaches can limit the ability of accountable entities to calculate measure scores on their own. This inability to replicate measure scores can limit the usability of risk-adjusted performance scores for local quality improvement efforts. It is also not meant to diminish the investigation into diseases and processes that need novel measure development. These recommendations and standards are intended to advance measurement science in numerous areas, such as identifying and testing data sources, and these standards will facilitate consistency in the evaluation of risk adjustment models within performance measures for NQF endorsement.

NQF endorses measures through multistakeholder Committees, which achieve consensus agreement on a set of standard **endorsement criteria**. This guidance will further advance NQF's measure evaluation criteria by informing future revisions to NQF's endorsement standards for social and functional risk adjustment. The agreed-upon best practices that developers should implement, at a minimum, are listed below:

Risk Adjustment Minimum Standards

1	A conceptual model is required and should illustrate the hypothesized pathways between the social and/or functional status-related risk factors, patient clinical factors, quality of care, and the measured healthcare outcome. At a minimum, the conceptual model must be supported by a literature review. Developers may also consider supplementing the literature review with expert opinion (e.g., clinical experts, patients).
2	Developers must, at a minimum, consider age, gender, race, ethnicity, an indicator of urbanicity/rurality, indicator of poverty (such as Medicare and Medicaid, dual eligibility, and income), indices of social vulnerability (such as the Area Deprivation Index [ADI] or the Agency for Healthcare Research and Quality [AHRQ] SES Index score), and indicators of frailty and disability (such as ADLs, vision, hearing, cognitive impairment, eligibility for disability programs, etc.) for inclusion during the process of developing the conceptual model.
3	If social and/or functional risk factor data are not available in a data source of sufficient quality (see the section titled **Identifying and Selecting Potential Data Sources and Variables**) but these factors were included in the conceptual model, then the developer should describe the potential bias that may exist, and its direction, as a result of not including the risk factor(s) in the risk adjustment model. The developer should also provide a justification for why the measure still has validity in this circumstance.
4	Developers should document and fully disclose data sources, including the dates of data collection, any data cleaning and manipulation, and the data's assumed quality. Developers can cite other research to show data quality of variables. Developers should also describe the populations covered within each data set.
5	Developers should provide descriptive statistics on how the risk variables identified from the conceptual model are distributed across the measured entities.
6	Model calibration should be conducted within the overall population and within relevant subgroups defined by clinical, social, and functional risk factors that may bias the outcome. Measures that include social or functional risk factors in the final risk model should be calibrated in subgroups defined by those factors to the extent possible. At a minimum, developers should conduct subgroup calibration analyses for race, ethnicity, an indicator of urbanicity/rurality, an indicator of poverty, and an indicator of disability. Developers should be transparent about their approach and their interpretation of the results.

7	To maximize the ability to identify healthcare disparities, the final measure specifications should provide a stratification approach for calculating and displaying measure scores by at-risk subgroups. At a minimum, developers should, to the extent feasible given the samples' sizes, use race, ethnicity, an indicator of urbanicity/rurality, an indicator of poverty, and an indicator of disability for subgroup stratification analyses. Beyond this minimum set of factors for subgroup stratification analysis, developers should consider stratification to distinguish between groups of patients who may have difficulties accessing care, for example. This should be informed by literature, patients, experts, or other stakeholders and is reflected in the conceptual model. The distribution of the measure scores and sample sizes for each subgroup across providers should also be presented.

Conceptualizing the Model

Developing the Conceptual Model

A conceptual model illustrates the hypothesized pathways between the social and/or functional risk factors, patient clinical factors, healthcare processes, and the measured healthcare outcome. The pathways between risk factors and the care process should be illustrated and accompanied by evidence of the relationships. A well-developed conceptual model is informed by clinical and population health research literature reviews and can be supplemented by expert opinion (e.g., clinical experts, patients). Each piece must be considered specifically and described in relation to the measured outcome (**Minimum Standard #1**).

1	A conceptual model is required and should illustrate the hypothesized pathways between the social and/or functional status-related risk factors, patient clinical factors, quality of care, and the measured healthcare outcome. At a minimum, the conceptual model must be supported by a literature review. Developers may also consider supplementing the literature review with expert opinion (e.g., clinical experts, patients).

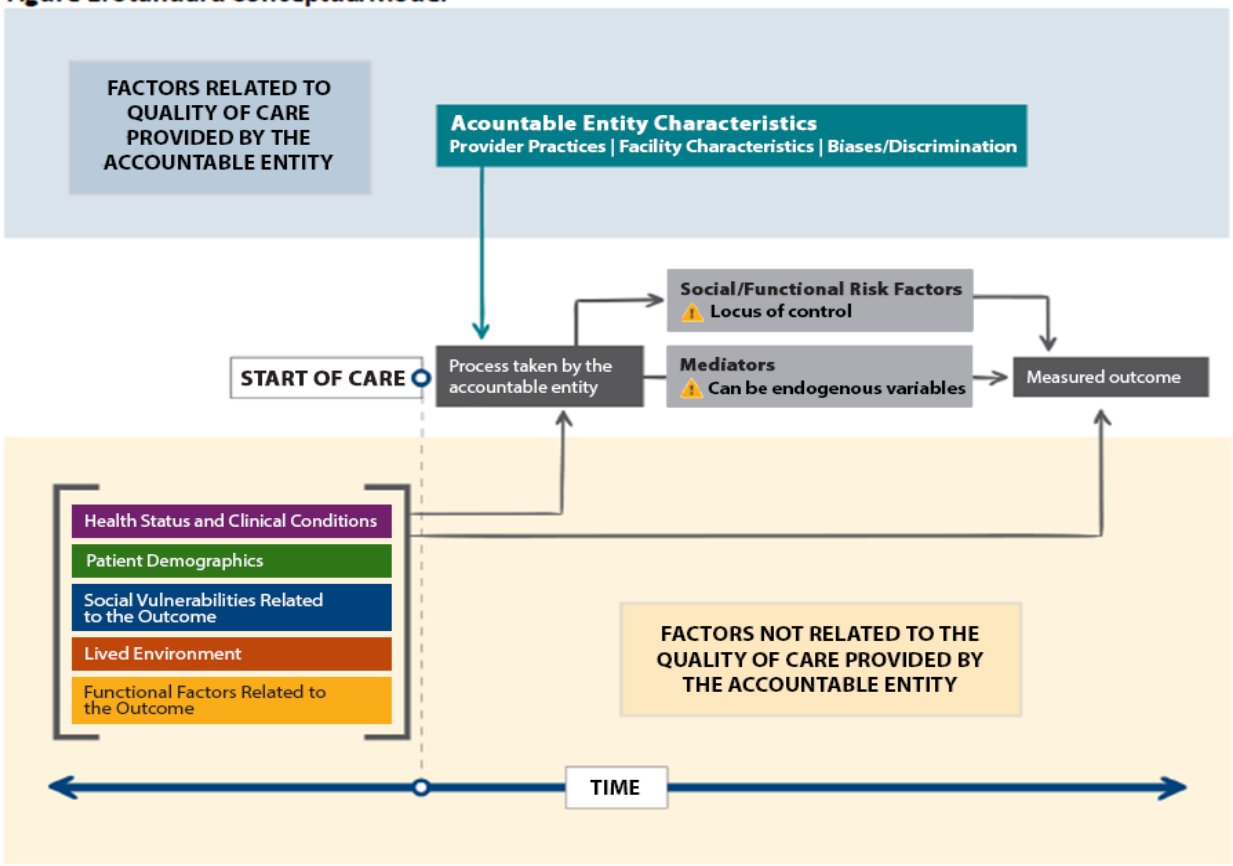
The conceptual model is only the first step in determining an appropriate risk adjustment model, but it is extremely important to engage in this work prior to empirically testing any risk factors. By mapping these relationships, measure developers can clearly communicate the evidence base and assumptions that will guide their decisions about the final risk adjustment model. The conceptual model can then be relied upon at later stages to make determinations about the importance or marginal effects of risk factors once empirical data are available.

Quality measures should adjust for patient characteristics at the start of care (i.e., measurement period) that affect the measured outcome but are not an indicator or characteristic of the care provided (e.g., treatments, expertise of staff). Therefore, all relevant demographic factors, clinical risk factors, social and functional risk factors, and patient preferences related to the outcome of interest should be considered for inclusion in the conceptual model. The conceptual model depicts all relevant factors, regardless of whether they will be used in the final risk adjustment model or whether data can be operationalized in the full measured population. What is essential at this step is to consider relevant factors for sufficient evidence of a relationship with the outcome and to depict them in the model. For example, accountable-entity characteristics, such as provider practice size, facility characteristics and resources, and potential for entity biases/discrimination, should be explored and mapped in the conceptual model because they help describe the care pathways targeted by the measure.

As described in **Minimum Standard #3** and in the **Identifying and Selecting Potential Data Sources and Variables** step, social and/or functional risk factors may be identified in the conceptual model; however, there may be data limitations present that will have an impact on their use as variables within the risk model. For example, housing instability can affect a patient's ability to store certain diabetes medications (e.g., insulin), which may result in their ineffectiveness and subsequently can contribute to poor health outcomes. Yet data to identify patients' housing status may not be available or may only exist at the community level rather than the patient level. Nevertheless, lack of data for an essential risk factor should not prohibit the consideration of these factors within the conceptual model. All influential factors should be identified in the conceptual model regardless of whether they can be operationalized in available data.

It is strongly recommended that developers construct a graphical representation of these relationships for clarity and ease of analysis by measure review Committees. Figure 1 provides a graphical depiction of a standard conceptual model. It depicts a template for developers to visualize the basic structure of a conceptual model. Developers can use and subsequently populate the template once all the evidence to support the aforementioned relationships is gathered. The remainder of this section describes considerations for identifying the contents of this conceptual model template. An example graphic from an NQF CDP measure submission is also presented with **Example 2** in **Appendix D**.

Figure 1: Standard Conceptual Model



CHARACTERISTICS OF THE PATIENT

In Figure 1, the timeline at the bottom shows the impact that time has on all other elements of the graphic. The graphic is also separated into a top portion, which includes factors related to the quality of care provided by accountable entities, and a bottom portion, which includes factors potentially influencing the measured outcome that are present at the start of care (i.e., patient characteristics present at the start of care). Starting on the left are the patient characteristics present at the start of care (in brackets). These characteristics, or factors, are important to identify and consider for risk adjustment because they are what the patient brings into the encounter at the "start of care" point (i.e., start of measurement period).

CHARACTERISTICS OF THE ACCOUNTABLE ENTITY

The teal-colored box shows the accountable entity's characteristics, which can include, but are not limited to, provider practices (e.g., adequate discharge planning), potential biases and/or discrimination based on patient characteristics (e.g., inability to provide or prioritize language services or culturally competent care, other -isms), and facility characteristics (e.g., safety net providers or critical access facilities). Because this box contains characteristics related to the quality of care provided by the accountable entity, and as such can be controlled by the accountable entity, it is placed in the top portion of the graphic. Those accountable-entity characteristics ***should not*** be used as variables to risk-adjust the measure of interest. Again, these should be included in the conceptual model, but the

measure should not be adjusted to account for differences across entities because accountable entities have a responsibility to understand the needs of the populations they serve and address risk factors that they can influence through changes in care delivery.

MAPPING THE PROCESSES OF CARE AND THE MEASURED OUTCOME

The middle portion of the graphic contains the hypothesized care pathways. Developers need to map this section so that they can determine the factors that are related and unrelated to the quality of care provided by the accountable entity. This mapping will directly influence the structure of the risk adjustment model.

We begin with the start of care and processes taken by the accountable entity. An arrow between the processes and the measured outcome represents the accountable entity's actions to achieve the measured outcome. The actions taken by the accountable entity will lead to the measured outcome, but in between those steps, mediators and other social/functional risk factors can have an impact on that outcome depending on the accountable entity's locus of control. Yellow caution signs denote that developers should carefully consider the risk factors in these boxes in particular, given the subsequent steps in this Technical Guidance.

The arrow from the patient characteristics to the processes of care and the measured outcome boxes illustrate that characteristics already present at the start of care can influence these elements of the care pathways. The arrow from the accountable entity's characteristics to the processes of care illustrates that the accountable entity's characteristics can influence the approaches it takes to caring for the patient and the specific care processes it provides.

AREAS OF SPECIAL CONSIDERATION

Social and functional risk factors should be reviewed and graphed separately from other mediators. When identifying the social and functional risk factors, measure developers should carefully assess the current locus of control of the accountable entity to meaningfully influence the risk factor and place boxes appropriately in the top or bottom of the graphic. For instance, in this graphic depiction, if it is within the provider's locus of control to affect the social and/or functional risk factor by taking certain actions (i.e., they can meaningfully influence and change the factor's effect on the outcome), then the risk factor would be placed in the top portion of the figure to reflect that it is associated with the quality of care provided. If the risk factor is not under the accountable entity's locus of control, it should be placed at the bottom of the figure to make clear its potential for inclusion in the risk adjustment model and that it is not influenced by the quality of care provided by the accountable entity. Because a risk factor is included in the conceptual model, this asserts that the risk factor should be considered by the accountable entity when determining a patient's treatment plan. For example, when adapting NQF #3597 in **Example 2** of **Appendix D** from an ACO level to a clinician level of measurement, the developer identified that clinicians participating in the Merit-Based Incentive Payment System (MIPS) may have the ability to mitigate some or all of the risk conferred by a patient's SES. For example, a provider can consider a patient's education level, health literacy level, and home living situation when planning and delivering care. In addition, high quality care may be characterized as being more racially, linguistically, and culturally sensitive and informed. Measure developers must examine whether and how much it is

within the accountable entity's locus of control to mitigate the risk factor. Developers can demonstrate an accountable entity's ability to meaningfully influence a factor by citing the primary literature, public reports, and case studies and/or by conducting empirical analyses (as described in the **Locus of Control** section) to determine the variation and degree of impact of a risk factor to a measured outcome. Variation suggests that some entities can successfully mitigate a risk factor, as reflected by observed differences in outcomes (i.e., that at least for a subset of entities, those factors can be impacted). The locus of control should not be presumed to be fixed or immutable based on current practice, guidelines, and capabilities, or to be identical across diseases, conditions, and entities. However, developers should also consider the impact that variation in the availability of community-based resources and state policies have on an accountable entity's ability to address social and/or functional risk, as these factors may vary across measured entities. This impact may not always be quantifiable, but developers should discuss these considerations as a limitation, if needed.

Separately, **mediator variables** may also explain the observed relationship between the actions of the accountable entity and the measured outcome, and therefore, they should be examined prior to their inclusion in the risk adjustment model as well. For example, unplanned hospital readmissions may be mediated by postoperative, surgical complications because these complications may exist in the causal pathways between the actions of the accountable entity and the measured outcome of readmissions (i.e., an entity's action may lead to a complication, which may cause an unplanned readmission). When identifying mediators, developers should be aware that **endogenous variables** can manifest as intermediate clinical outcomes, and intermediate clinical outcomes ***should not*** be adjusted for, considering that they lie along the care pathways and relate to the quality of care of the accountable entity. In other words, if providing a specific treatment is critical to the care process, then this action should not be a risk factor in a risk model.

REITERATING THE PURPOSE OF THE CONCEPTUAL MODEL FIGURE

The conceptual model serves as the foundation for the remaining steps outlined in this Technical Guidance. The risk adjustment model can be misleading and ineffective unless it is grounded in a transparent conceptual model informed by literature review, which can be supplemented by expert input (e.g., clinicians, patients).

Measure developers should be mindful that issues of access to care affect the conceptual model, namely that having access to care is, generally, a prerequisite for inclusion of patient characteristic information within the data sources used to develop risk adjustment models. While we often do not have data on patients who do not or cannot access care, accountable entities may seek to improve their access to high quality care. Risk adjustment alone cannot solve this healthcare access issue. Instead, other measurement and structural approaches (e.g., developing a balancing measure to capture patient populations with healthcare access issues with different data, or promoting policies that reduce Medicaid enrollment churn) can be used to help address issues of access. However, the TEP recommended that developers disclose, at the conceptual model step, whether there are patients who are not captured in the data sources used, and subsequently within the measure, due to healthcare access issues.

The conceptual model will also inform decisions about stratification of the measure score by relevant at-risk subgroups (see the section titled **Stratification** for more information). These decisions relate to the design of the conceptual model and the measure's specifications; therefore, they should be outlined in a narrative description of the model as well. This will help others who are uninvolved in the conceptual model's development to understand what decisions were made and why. At a minimum, the TEP recommended that race, ethnicity, an indicator of urbanicity/rurality, an indicator of poverty, and an indicator of disability be considered for subgroup stratification analysis for all risk-adjusted measures. Therefore, these factors must also be considered in the conceptual model and overlap with the factors described in **Minimum Standard #2**.

In all, developers should write a brief description of their processes for developing the conceptual model using citations to establish the relationship between factors and outcomes, explaining when literature or data are unavailable to examine a relationship, and reporting how expert opinion was incorporated into the decision-making processes, if used.

Factor Selection for Inclusion in the Conceptual Model

For the first step, measure developers should explore the broad list of factors that might have an impact on the outcome. This Technical Guidance is not overly prescriptive about requirements for evidence so as not to burden measure developers or hinder development in unmeasured areas of disease and healthcare. At a minimum, evidence for inclusion in the conceptual model should include a literature review to identify the most important and plausible factors. A systematic review is not required, and it may be advisable to turn to the grey literature for evidence as well. Developers should also attempt to supplement the literature review with expert opinion to identify and uncover risk factors not found in the literature. Although they can be useful and are encouraged, internal empirical analyses are not required at this stage. However, developers should be sure to focus on finding evidence of variation in care to demonstrate the importance of the risk factor and its effect on the measured outcome. For example, clinical TEPs are often convened to identify a list of functional risk factors associated with the outcome of interest via a modified Delphi method or nominal group technique.[39] Measure developers will also look to public health, sociological, and medical literature for investigations into the impact of social and/or functional risk factors on measured health outcomes. The patient community (e.g., patients, caregivers, and patient advocates) may also be involved as experts in order to verify or further examine the influence these risk factors can have on the measured outcome, as this can reveal additional factors for consideration or explain a potential confounding relationship. Developers and experts may anticipate that some factors may be duplicative or exert the same level of influence on the outcome, and thus, they should not be included in the final risk adjustment model. However, these factors should all be reflected in the conceptual model. They could be eliminated during the testing phase when developers are able to identify statistical issues (e.g., overfitting, multicollinearity, and/or confounding) in the model's structure, which may lead to decisions to exclude certain factors in order to remove biases from the model. Once the conceptual model is fully drafted, developers should review their results from end to start (i.e., right to left in Figure 1). Moving backwards through the model can help to identify assumptions that were made or logical fallacies that may otherwise go unnoticed.[40]

When designing the conceptual model, it is important to remember that these factors can have either a direct or indirect effect (i.e., via the actions taken by the accountable entity) on the measured outcome.[41] Both the direct and indirect effects of factors should be considered for model inclusion. Endogenous (i.e., dependent) factors other than the outcome of interest should be identified in the conceptual model because they are also associated/vary with the outcome of interest. However, endogenous variables should be used with caution in the final risk adjustment model, as they may raise the potential for biased results. For example, these endogenous variables could manifest as intermediate clinical outcomes, such as complications that reflect the quality of care and increase risk of the measured outcome. It is not appropriate to adjust for those variables that were not present at the start of care.

Developers may find that it would be more accurate to combine several risk factors into a construct for the model. For example, a social risk factor of low social support could be characterized as a construct of three variables: (1) marital status, (2) living alone, and (3) utilizing home health aide support. This is also true for functional risk factors. For example, a construct for frailty could include three variables relevant to the measured outcome: (1) use of walkers, (2) use of oxygen, and (3) receiving disability insurance benefits.

Similarly, measure developers need to evaluate evidence of whether the social and/or functional risk factor has little or no influence on the outcome. Both inclusion in and exclusion from the conceptual model should be mindfully considered, especially for factors that must be considered as described in the paragraph below and for which disagreement exists regarding their effect and the measured entity's ability to influence it.

There are a number of social and functional risk variables that must always be considered in the process of developing the conceptual model for outcome and cost/resource measures. This best practice was agreed upon by the TEP, which identified a minimum set of factors that are commonly used and analyzed by developers. The TEP further determined that data to support analyses of these factors as variables in the risk model are largely available, reliable, valid, and generalizable. The following set of factors must be considered, at a minimum, for examination in the process of developing conceptual models: age, gender, race, ethnicity, an indicator of urbanicity/rurality, indicator of poverty (such as Medicare and Medicaid dual eligibility and income), indices of social vulnerability (such as the ADI or the AHRQ SES Index score), and indicators of frailty and disability (such as ADLs, vision, hearing, cognitive impairment, eligibility for disability programs, etc.) (**Minimum Standard #2**). The consideration of these factors within the conceptual model ***is not*** a requirement for their use in the final risk adjustment model because this is dependent upon their relationship to the outcome of interest. This list is a minimum standard and is not meant to limit measure developers from exploring other social and functional risk factors that are relevant to the specific measure focus. For example, if relevant, the developer must specify whether biological sex versus self-identified gender is intended, as this guidance is flexible to accommodate either definition. However, developers should describe the rationale for including or excluding all factors in the final risk adjustment model that were considered in the conceptual model, including the above minimum standard list.

2	Developers must, at a minimum, consider age, gender, race, ethnicity, an indicator of urbanicity/rurality, indicator of poverty (such as Medicare and Medicaid dual eligibility and income), indices of social vulnerability (such as the ADI or the AHRQ SES Index score), and indicators of frailty and disability (such as ADLs, vision, hearing, cognitive impairment, eligibility for disability programs, etc.) for inclusion during the process of developing the conceptual model.

EXAMINATION OF RACE FOR INCLUSION IN THE CONCEPTUAL MODEL

Generally, decision-making approaches with respect to risk factor selection for the conceptual model, and subsequently the variable selection for the risk model (see section titled **Operationalizing Variables for Risk Adjustment**), should be consistently applied for any risk factor or variable (i.e., clinical, social, or functional). However, there may be certain risk factors and variables, such as race, for which the decision to either include or exclude these elements requires additional consideration, including the effect of such variables on health equity. The focus of this section is to inform the measure developer's decision-making process for considering race as a risk factor in the conceptual model. Decisions at this stage will influence subsequent risk variable decisions for the final risk adjustment model. Additional guidance for variable selection and empirical testing of the risk model can also be found in later sections of this guidance.

In reviewing the stakeholder feedback (**Appendix F**) and sharing its perspectives and insights from the literature, the TEP found inconsistent and often conflicting perspectives regarding the inclusion or exclusion of race variables. Although there are a variety of rationales that have been proposed for including or not including race in risk adjustment models, the TEP came to an impasse on a recommendation for the inclusion or the exclusion of race, completely, from the conceptual model and subsequent risk adjustment models. This section includes the considerations from stakeholders and TEP members that the developer should consider with respect to race.

The TEP acknowledged that there are a variety of rationales, or mechanisms, that have been historically proposed for including race in healthcare risk adjustment models: (1) to account for genetically mediated differences in disease predisposition that may be associated with race, risk of clinical outcomes, or response to treatment (e.g., certain medications); (2) to account for the effects of discrimination based on race (i.e., racism); and (3) to use as a proxy for other social structures, effects, and residual differences not fully reflected by other factors. The first mechanism can occur when a measure developer seeks to adjust for evidence-based differences in biological factors that vary by race and that affect the outcome. For example, evidence exists that shows greater risk for some diseases (e.g., Apolipoprotein L1 and chronic kidney disease) in Black patients with certain gene variants.[42] The TEP generally agreed that this may be an appropriate use of race in the conceptual model if the risk factor cannot be represented directly. However, the TEP did not come to a consensus on the remaining mechanisms.

The TEP considered and voiced a number of arguments against adjusting for race for the latter two mechanisms. Feedback from the stakeholder focus groups during the second phase of this work

expressed caution with the use of race, noting that race should not be used as a risk adjustment variable because it is unclear what the variable truly reflects, and its use might lead to reduced transparency or inaccuracy of risk models. Some TEP members agreed that race, as a risk variable, could reflect interpersonal and structural racism. Other TEP members and stakeholders (**Appendix F**) also noted that race was an inadequate proxy to be used in a conceptual model, suggesting that measure developers should identify which variables are truly explanatory in their conceptual model and not rely on the social construct of race as a proxy for other unmeasured factors; some stakeholders commented that this approach to using race could perpetuate the misconception that social needs and social risks are invariably connected to race. Additionally, conflating concepts of race and risk without understanding the precise mechanisms that impart the risk can lead to harm. For example, using race as a risk-adjuster based on statistical associations in historical data, which would generally show worse outcomes for racial minorities, has the potential to inadvertently perpetuate long-standing disparities by setting differing expected outcomes for previously disadvantaged races.

Other TEP members called for more flexibility in the guidance on the use of race in risk models. In situations where only race data are available but other more specific variables (e.g., granular social risk data; detailed, personalized genetic information) are not, the inclusion of a race variable, although imperfect, may be the best available proxy for unmeasured social risk factors. Additionally, since there are underlying mechanisms associated with race that are poorly understood (i.e., because unknown exogenous factors impact the risk model), the use of race may facilitate capture and better understanding of those latent concepts and improve the risk model performance. They voiced concern that the threat of perceived risk aversion (i.e., the perceived risk of not serving certain populations that may reduce the overall measure score if risk adjustment is not applied) by accountable entities is too substantial to broadly and indiscriminately advise against adjustment for race. They expressed that prohibition of adjustment for race in risk models could worsen rather than mitigate health inequities by reducing access to care for minorities. Lastly, risk adjustment models that do not adjust for race might have poor calibration and inaccurate risk prediction, especially within some minority populations.

Ultimately, given these cogent but conflicting arguments on both sides of this issue, TEP members agreed that race should be considered carefully on a case-by-case basis for each measure, with special consideration to the measure's **intent and use**, if known. Measure developers must disclose the rationale for including or excluding race, which should take into consideration the underlying mechanisms associated with race as described above.

Locus of Control

Within the conceptual model, it should be clear which steps and processes the accountable entities can influence to improve the measured outcome and those which they cannot influence. Considerations might include the degree of control that accountable entities have to affect or meaningfully influence outcomes, which may vary by context.[36] For example, developers should consider the variability in the availability of community-based resources, provider availability, and state policies, which may affect accountable entities' options and abilities. As explained previously, evidence to support these decisions can similarly be taken from a combination of sources, such as expert opinions, literature review of peer-reviewed articles and white papers, and/or from conducting 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 in light of the locus of control considered. For example, when adapting **NQF #1789** (see **Example 3** in **Appendix D**) from a hospital level to an ACO level of measurement, the measure developer constructed a conceptual model outlining the relationships between the potential, clinical, and contextual factors and rates of readmission at the ACO level. These contextual factors included the following: (1) physical environment (e.g., green spaces, safe streets); (2) community resources (e.g., home health, senior services); (3) patient resources (e.g., social support, transportation, and income); and (4) patient behavior/personal preferences (e.g., exercise, diet, advanced care directives, and preference for intervention). The developer noted that these contextual factors, which are different than traditional medical care delivered in the hospital settings, will have an impact on the likelihood of readmission. The developer identified that ACOs practicing in communities where patients have limited access to transportation, healthy foods, and recreational facilities may have less success in promoting healthy behaviors among patients. This may, in turn, have an impact on readmission rates. Thus, the conceptual model recognized the capacity of ACOs to mitigate the effects of many contextual factors on rates of admissions, encompassing both SES and non-SES variables. Therefore, the same measure developed for hospitals, which was adapted for ACOs, required different risk adjustment models due to their differing loci of control.

Furthermore, the conceptual model should consider whether it is feasible for accountable entities to diminish the impact of social or functional risk factors, as measurement should be accurate and reflect the agreed-upon scope of responsibility of the accountable entity as opposed to being too aspirational in expanding the locus of control of the measured entity beyond its achievable scope. Developers should consider evidence that demonstrates that accountable entities can or cannot mitigate the effect of the social or functional risk factors linked to the measured outcome. It can be difficult to quantify all of these impacts on the measured outcome, but developers should state this in the discussion of their model.

Intended Use

The developer must examine the role of social and/or functional risk factors in the context of the specific intended use of the measure, including the locus of control and the structure of the program that will use it, if known. The specific intended use of the measure may include public reporting; payment applications, such as VBP, shared savings programs, or other risk-bearing arrangements; quality improvement; or other policy and research applications.

Measure developers must consider the intended use of the measure from several perspectives and the level of measurement (e.g., hospital, health plan, ACO, and clinician) as described in the section above. The developer should explain, to the extent possible, how the intended use affects the developer's choices regarding risk variable inclusion, feasibility of stratification, and potential unintended consequences. Additionally, measure developers are and should be required to identify and discuss the potential for unintended consequences of the measure due to the measure's use. The benefits of measure use must outweigh the potential negative unintended consequences of measurement (see the section titled **Negative Unintended Consequences**).

For example, measures tied to strong financial incentives that are used for VBP should consider the evidence regarding how accountable entities can take specific actions to ameliorate the impact of social and/or functional risk to the measured outcome. In VBP scenarios, it is important to reduce the potential for risk aversion, especially for some accountable entities (e.g., safety net providers) who serve a disproportionate number of patients with social and/or functional risk factors and have concerns about disproportionate penalties.[28] The conceptual model should outline the evidence demonstrating the ability of accountable entities to positively influence outcomes through mitigating social risk factors (i.e., literature review, expert opinion, and examples of what accountable entities with better outcomes compared to other measured entities have done to achieve better outcomes). Developers should re-evaluate social and/or functional risk adjustment when adapting measures for other uses, given this relationship between the measure's use and a measured accountable entity's ability to mitigate social risk factors (i.e., locus of control).

Measure developers can look to **Appendix D** for an example of verbatim text from an NQF-endorsed measure submission. **Example 1** depicts a developer's explanation of a conceptual model, including the literature review and caveats up front for the hypotheses made during the conceptualization of the model. **Example 2** shows a conceptual model graphic accompanied by a narrative description of its contents. Please note that both NQF-endorsed measures were submitted prior to the release of this Technical Guidance and will not reflect all of the best practices described in this document. However, they are provided as illustrative examples of developing a conceptual model.

Identifying and Selecting Potential Data Sources and Variables

Once social and/or functional risk factors are identified within the conceptual model, 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 present that will impact their use as variables within the risk model. 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 due to the inability to account for potential confounding. The developer should also provide a justification for why the measure still has validity even in this circumstance.

Risk adjustment increases the likelihood of fair comparisons of performance across accountable entities. It accomplishes this by controlling for confounders. Confounding factors may be clinical, social, and functional in nature. If confounding factors are not adequately controlled for, the risk model may result in spurious and inaccurate estimates of performance due to this confounding bias.

At a minimum, developers should examine previously published evidence and should attempt to estimate the directionality of the bias for the factor of interest by using other studies. This may be achieved by reviewing the literature to determine how a risk factor might affect subsets of the accountable entities due to how those patients who are at risk to the outcome (due to the factor of interest) are distributed across the accountable entities. For example, if frailty has been identified in the conceptual model but cannot be identified in a data source of sufficient quality (see Table 1) for

inclusion as a variable in the risk model, the developer may review the literature to determine the degree of bias that may exist by showing that the factor is not evenly distributed across the accountable entities of interest (**Minimum Standard #3**). This uneven distribution may result in a bias for a subset of accountable entities that disproportionately serve patients with the social risk factor. The developer may also consider including proxy variables of social and/or functional risk in the risk adjustment model based on prior research. This approach could reduce the impact of bias on the measure by accounting for some, but not all, of the effect that the otherwise missing variable would have on the model. The relevance of these proxy variables should be conceptually and empirically appropriate for the measured outcome of interest (see the section titled **Operationalizing Variables for Risk Adjustment** for more information about the use of proxy variables).

The TEP acknowledged the substantial trade-offs in approaches to addressing bias. For example, if there is a high degree of difference across entities in a social risk factor that cannot be accounted for, the developer may recommend excluding entities that serve a disproportionately large number of individuals with that risk factor upon application and when data are available. However, this strategy should be considered judiciously and noted clearly due to the potential downside of excluding measured entities that serve patients for which there is a need to advance quality and health equity, such as the exclusion of key providers, such as Federally Qualified Health Centers.

---

3 If social and/or functional risk factor data are not available in a data source of sufficient quality (Table 1) but these factors were included in the conceptual model, then the developer should describe the potential bias that may exist, and its direction, as a result of not including the risk factor(s) in the risk adjustment model. The developer should also provide a justification for why the measure still has validity in this circumstance.

Transparency is one of the core principles of risk adjustment. Therefore, developers should explain their rationale for using selected data sources or samples and offer justification for the data's appropriateness. Developers must ensure these data are reliable, valid, complete, comprehensive, timely, and generalizable (see Table 1). Therefore, the developer should document and fully disclose the data sources used, including the dates of data collection; the manner of data cleaning and manipulation, if done; and the data's quality (e.g., whether the data are periodically audited) (**Minimum Standard #4**). Developers should also provide a description of the populations covered within that data set (e.g., all age groups and payers, or limited to 65 years of age or older, or limited to those on Medicare).

4	Developers should document and fully disclose data sources, including the dates of data collection, any data cleaning or manipulation, and the data's assumed quality. Developers can cite other research to show data quality of variables. Developers should also describe the populations covered within each data set.

Table 1. Considerations for Assessing Data Quality

Consideration	Description
Reliable	The method of collection must be reproducible with minimal variation between one collection and another if the same population is the source.
Valid	Validation ultimately rests on the strength of the logical connection between the construct of interest and the results of operationalizing their measurement, recording, storage, and retrieval.
Complete	Data should contain as few missing values as possible, and the allowable percent missingness should be stated. Missing values are difficult to interpret, and they lower the validity of the model. Missingness should be evaluated as to cause (e.g., the Rubin taxonomy, which includes missing completely at random; missing at random; or missing not at random).
Comprehensive	Data are sufficiently comprehensive to adjust for known and suspected risk factors in the causal model and to limit the number of proxy measures required for the model. Obtaining the primary information is sometimes impossible; therefore, some proxy measures might be inevitable for certain projects.
Timely	Data are as recent as possible. If the measure developer used 1990 data in a model designed for use in 2021, many people would argue that the healthcare system has changed so much since 1990 that the model may not be relevant.
Generalizable	Steps to ensure findings can be generalized to target populations should also be taken when developing the model. Findings from algorithms based on populations of limited size and scope should be validated in broader populations to ensure generalizability.

Risk adjustment of outcome measures, including cost/resource use, includes statistical procedures that rely on sufficient sample size to produce reliable risk estimates. When creating a risk adjustment model, there should be sufficient data available to ensure a valid model (see the section titled **Empirically Testing the Adequacy of the Risk Model**).

Different statistical reliability estimates apply to different types of models. For example, a measure with an outcome that is more common may require fewer cases per accountable entity to consistently return an acceptable level of reliability across measured providers. If the outcome is uncommon, then the number of cases required could be much larger.[14] Other factors may also affect the size needed for a sample, such as a lack of variability among risk factors for a small sample that results in partial correlation (also known as *collinearity*) among risk factors and a corresponding decrease in the stability of the parameter estimates (i.e., when predictor variables in the same regression model are correlated, they cannot independently predict the value of the dependent variable). In general, the larger the sample size is, the greater the statistical power to detect outliers is and the higher the measure reliability will be.

Measure developers can look to **Appendix D** for an example of verbatim text from an NQF-endorsed measure submission. **Example 3** depicts how a developer examined the literature to identify social risk factors for the conceptual model and subsequently used the conceptual model to guide the selection of variables (both clinical and social) for empirical examination. This measure example exhibits how a developer examined a conceptual model at multiple loci of control. Of note, the NQF-endorsed measure was submitted to NQF prior to the release of this Technical Guidance.

Common and Emerging Data Sources

Data for social and/or functional status risk adjustment within quality performance measures can come from a variety of sources, each with respective strengths and limitations depending on the measure context (**Appendix C**). The most frequently used data sources are administrative claims data, registry data, clinical assessments, patient-reported surveys/instruments, and EHRs.[38] Of these, the most common data source for developing risk adjustment models is claims data, namely Medicare fee-for-service (FFS) claims, and clinical registry data.

However, novel and emerging data sources may also be of use, noting the data quality considerations mentioned previously (Table 1). Recent developments in data standardization may help with data availability for more accurate measurement of and adjustment for social and/or functional risk factors. For instance, the Robert Wood Johnson Foundation (RWJF)-sponsored Gravity Project is creating standardized items and tools using the Health Level Seven (HL7) Fast Health Interoperability Resource (FHIR) to more uniformly collect data on SDOH, such as housing, food security, and transportation.[43] Similarly, the CMS-sponsored PACIO Project is developing item sets for cognitive impairment and frailty, areas of functional status that have had ambiguous definitions and scarce data.[44]

Additional sources for information on social risk factors could include the International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM) Z codes, which identify nonmedical factors that may influence a person's health status. Existing Z codes identify issues related to a patient's

socioeconomic situation, including education and literacy; employment; housing; lack of adequate food or water; or occupational exposure to risk factors, such as dust, radiation, or toxic agents.[45] However, Z codes are currently in very limited use within claims.[45] Using Z codes within risk adjustment models will only be practical if providers adopt these codes more widely. Social risk information may also be collected from standardized assessment tools, such as the Protocol for Responding to and Assessing Patients' Assets, Risks, and Experiences (PRAPARE) assessment tool, which collects SDOH data across the national network of federally qualified health centers and Medicaid-managed care organizations.[46] Developers may also consider the potential contribution of indirect estimation methods, which seek to derive demographic parameters from indicators that are largely, but not entirely, determined by the specific parameter of interest. For instance, geographic assignment methods based on the United States (U.S.) Census[47], the ACS Data[48], the ADI[49], or the Bayesian Indirect Surname Geocoding[50] may be used to support the identification of social risk factors. However, developers should use caution; the data used should be reviewed for accuracy and bias as the U.S. population becomes more diverse.

Once data sources are identified and permissions are arranged (i.e., **data use agreements**), relevant databases may need to be linked and various data preparation tasks performed, including an assessment of the data reliability and validity, if not previously confirmed. If samples are being used, the measure developer should draw them using predefined criteria and methodologically sound sampling techniques. Testing to determine the suitability of data sources and differences across data sources may also be necessary.

Empirically Testing Risk Factors

After examining the data sources and variables available to capture these identified risk factors, developers should empirically test the social and/or functional risk factors. When a risk factor has been identified as exogenous and appropriate in the conceptual model, then using statistical significance testing for social or functional risk factor variables should not be deterministic for including or excluding that factor within the final risk adjustment model. The statistical cost of including an exogenous social and/or functional risk factor that is conceptually important, but without clear bivariable or multivariable significance, in the final risk adjustment model is minimal. Substantial variance inflation would indicate correlation with other adjustors or with reporting unit indicators in a regression model that predicts outcomes from adjustors and unit indicators in person-level data. Any increased variance needs to be balanced with reducing bias. The goal of risk adjustment is to accurately adjust for fair comparison rather than predict or clearly distinguish the conceptually appropriate risk factors that are uniquely responsible for specific aspects of adjustment. In terms of total (mean squared) error, even relatively small reductions in bias overwhelm potential variance inflation when sample sizes are large enough that precision is adequate, which is generally the case for NQF-endorsed measures. The rationale to exclude certain social and functional risk factors from the final model might include endogeneity. If a risk factor identified in the conceptual model is not included in the final risk adjustment model, the developer should, at a minimum, provide evidence that its removal does not bias the measure results for subgroups of patients. In addition, the factor may not be included if it imposes significant additional burden on collection and use.

The intent of this guidance is not to be prescriptive regarding the types of empirical testing that the developer should conduct. Empirical testing for social and/or functional risk factors is generally similar for clinical factors and may include an assessment of the relative effects of social and/or functional risk on measure performance and among subgroups of interest. Although the empirical testing is not deterministic, developers should examine that evidence in conjunction with the conceptual model. Developers should be transparent about their approach by describing the statistical methods used and the results and interpretation of the analyses, all of which lead to the decision on whether to select social and/or functional risk factors for risk adjustment.

Assessing the Variation in Prevalence of the Factor Across Measured Entities (i.e., Descriptive Statistics, Reporting Degree of Missingness of Factors)

At a minimum, developers should provide descriptive statistics on how the risk variables identified from the conceptual model are distributed across the measured (i.e., accountable) entities (**Minimum Standard #5**). Absolute or relative frequency statistics are examples of descriptive statistics that can be used for discrete social and/or functional risk factors.[51] This step should also examine any systematic missingness of variable collection across the measured entities.

This variation analysis is intended to describe the relationship between the risk factors and the measured entities. Variables with little variation in frequency across measured entities are not likely to be of statistical value in modeling performance differences across accountable entities, even if these factors have a significant association with outcomes. However, the analysis is intended to inform but not determine whether the factor is appropriate for inclusion in the risk adjustment model. For example, risk factors are often retained in the model in response to face validity testing.

---

5 Developers should provide descriptive statistics on how the risk variables identified from the conceptual model are distributed across the measured entities.

Operationalizing Variables for Risk Adjustment

Once the social and/or functional risk factors have been identified within the conceptual model, developers must then contemplate how to operationalize those factors into variables for inclusion in the risk adjustment model.

Variables meant to capture social or functional risk factors need careful consideration. For example, developers may determine in the conceptual model that a patient's income, as a social risk factor, has an impact on the outcome of interest. Patient-level data are preferrable for adjustment. However, due to data availability, operationalizing the patient-level social risk factor of income may only be performed using an **area-level variable** for income (e.g., county-level income) rather than at the individual patient level. This may not be sufficiently granular or specific, but due to data availability challenges, the area-level variable may be an appropriate proxy variable. Regardless, there can be instances in which area-

level factors are preferable, depending on what has been identified in the conceptual model. It is important to examine these factors in the model, as developers can then explain their logic behind selecting area-level variables or other types of proxy variables in substitution. When considering the use of area-level variables, developers should consider the heterogeneity of the population to ensure inferences made about the individuals within the geographic units can be generalized. Developers should also consider the subgroups within which they will test the calibration of the model, as mentioned later in this guide (see the section titled **Empirically Testing the Adequacy of the Risk Model**), and make clear in the conceptual model the reasons why subgroups may be affected by certain risk factors differently.

With respect to using proxy variables, developers should carefully consider the use of proxy factors when conceptualizing the model. Proxy variables should be based on prior empirical research and should be empirically appropriate for the intent of the measure. For example, geographic location may serve as a proxy for SES, provided that there is prior evidence that establishes a correlation between a geographic location factor and SES. It is vital that developers provide a clear explanation of the relationship between the proxy factor and the unmeasured social or functional risk concept. The potential unintended consequences of proxy variable use should be articulated. In particular, race requires careful consideration if used as a proxy variable, as discussed elsewhere (see the section titled **Examination of Race for Inclusion in the Conceptual Model**). Specifically, there is a lack of consensus on whether it is appropriate to use race to reflect, conceptually, other unidentified exogenous factors that can impact a risk adjustment model. When selecting variables, measure developers should consider and disclose the rationale for including or excluding race, which should take into consideration the underlying mechanisms developers assume race reflects.

Empirically Testing the Adequacy of the Risk Model

Measure developers should assess a risk model's performance by assessing its predictive ability, discrimination, and calibration. Each of these three methods will be examined individually below. Overall, to test the adequacy of a risk adjustment model, developers should describe the steps and methods of testing and the results of analyses used to validate the risk model performance. Candidate measure endorsement submissions should provide statistical results from testing the approach to control for differences in patient subgroups. These methods should be applied to test any risk-adjusted measure, not only those adjusting for social and/or functional risk.

There are various methods of assessing the predictive ability of a risk adjustment model. One approach is using statistics, such as explained variation (e.g., R-squared statistics), to quantify how close expected predictions are to the observed outcome. For the purposes of performance measure development, complete prediction in risk adjustment models is not always the goal. Rather, the goal is to adjust for differences in patient factors that differ across providers and are unrelated to quality to isolate differences in quality. Hence, the overall predictive ability of the risk adjustment model used in an outcome measure may be lower than that of clinical predictive risk models that estimate the risk of the same outcome.

Risk model discrimination is a critical model performance metric that calculates whether randomly selected patients who experience a particular outcome have a higher expected risk than those who do not experience the outcome. This can be quantified with separate measures of sensitivity or specificity, or more commonly using the c-statistic or Area Under the Receiver Operating Characteristic curve (AUC or c-statistic).[52] When considering the contribution of social risk and/or functional risk factors in modeling decisions, developers may compare the change in discrimination performance, such as AUC , with and without inclusion of social and/or functional risk factors and models that include clinical factors only. However, this analysis is not likely to be determinative. Improvement in the AUC may be limited when social and/or functional risk factors considered important in the conceptual model are added, especially if the standard, clinical factor-only model has a large baseline AUC.[53] Changes in model discrimination, such as c-statistics, can inform but should not solely determine decisions on whether to include an additional social and/or functional risk factor in the model specification.[54] Another approach builds on the work of Pencina and colleagues in evaluating the effect of an added predictor variable using integrated discrimination improvement (IDI) and the net reclassification improvement (NRI) approaches.[55–60] Measure developers should use these methods judiciously due to measure score sensitivity to choice and the number of category cutoff values, misleading results when models are poorly calibrated, and inflated assessments of added prognostic value.

Risk adjustment model performance must also be assessed to make sure the model is well calibrated across subgroups of patients (i.e., the relationship the model estimates between the risk factors and the outcome holds across patient groups and adequately predicts each group's risk, not just the risk of the patient population as a whole). Developers, for example, often test whether calibration holds in patients with the most advanced disease in disease-based measure cohorts. To adequately assess the impact of social and/or functional risk, risk adjustment model calibration must be examined within key subgroups in which the relationships of risk factors to outcomes might vary from that of the group as a whole. The TEP recommended that a minimum set of social and functional risk variables be used to test subgroup calibration, including race, ethnicity, an indicator of urbanicity/rurality, an indicator of poverty, and an indicator of disability (**Minimum Standard #6**). Beyond this minimum set of variables for subgroup calibration analysis, developers should assess calibration of any social and/or functional subgroups identified in the conceptual model specific to the measure. This minimum set of variables for subgroup analysis and the groups identified in the conceptual model are dependent on available data and sample sizes within subgroups.

Assessment of the adequacy of risk model calibration can be approached using various methods. Risk model calibration statistics inform whether the risk model-predicted probabilities are, on average, close to the average observed probabilities. The Hosmer-Lemeshow statistic is a commonly used approach to test statistical risk model calibration. However, this test is very sensitive to sample size (i.e., with large enough samples, the test statistic will always be significant).[61]

Graphical approaches to calibration are often preferred (e.g., plots of O/E outcomes across a broad range of expected values, such as patients grouped by deciles of expected risk). However, the interpretation of graphical approaches may depend on the measure, data availability, and sample sizes for subgroups, and adequacy of calibration requires subjective judgment. Optimally, model calibration

would be assessed in subgroups using models that include and exclude the SDS/SES variables being considered for incorporation in the risk model. If a model is intended to be used across a wide range of population subgroups and has very poor calibration in some of those subgroups (e.g., among rural subgroups), then the measure developer should make this transparent in the candidate measure endorsement documentation and any performance results derived from the measure. Caution should be used in the application of the measure score results for those poorly calibrated subgroups or for accountable entities that disproportionately care for patients from those subgroups. Poorly calibrated models may produce measure results that would be misleading to accountable entities who rely on these measures to target the healthcare needs for specific patient subgroups.

NQF recognizes that there is always tension between an overly narrow risk model with small sample sizes and restricted applicability, which only fits a very specific population, versus broader, all-inclusive, and more generalizable models with large sample sizes but whose calibration may not be as good for certain subgroups. There may be statistical or practical reasons for considering a separate model for certain population subgroups. For example, if there are data suggesting that one SES/SDS group has significantly better or worse outcomes than another or that the association of other variables with outcomes is considerably different for that subgroup, then it could be argued that they would be better served by a different model. Generally, the expected and observed values are close to being identical in a well-calibrated model. However, rare events (e.g., some never events) or small sample sizes for a subgroup pose challenges for these types of comparisons and may not be feasible. When a separate risk model is created for certain population subgroups, the risk model may be better calibrated than a risk model developed for the full population. However, this improvement in calibration from the individual population subgroup models comes at the cost of losing generalizability of the risk model to other populations, or to the full population. In these situations, developers should revisit the conceptual model to ensure that the goal of the measure for closing a particular quality gap is maintained. Developers may also consider excluding patients whose risk cannot adequately be predicted from the model or looking for more accurate risk adjusters for patients at highest risk (e.g., use of mechanical ventricular assist devices in patients with advanced heart failure to estimate their risk of mortality). Other reasons for a separate model might depend on the policy goals of the measure developer and measure implementer. The conceptual model can help developers to think about testing certain subgroups identified earlier in the development process. Developers should use caution when building separate models for subgroups unless there is sufficient sample size.

Lastly, risk models typically need to be periodically (re)calibrated to reflect the specific intended use and the characteristics of the measured population and temporal changes. When deciding whether to recalibrate, developers should consider the extent of the differences between how the model was originally developed and how it will be used.[46] For example, CMS adopted the hospital-wide readmission quality measure (**NQF #1789**), which predicted expected outcome rates for hospitalized patients and assigned accountability for patients to hospitals, for use in the **Medicare Shared Savings Program (MSSP) ACO program**, which assigns accountability for hospitalized patients to ACOs. The measure developer used a common set of variables across the risk adjustment models in each measure (i.e., the hospital-level measure and the ACO-level measures). However, given that the national set of hospitalized patients overlapped but differed somewhat across the two programs, the coefficients

associated with each variable in the risk models were re-estimated in the ACO data set and allowed to vary across the cohorts within each measure.[62] Testing results showed that the ACO model demonstrated adequate calibration.

---

6 Model calibration should be conducted within the overall population and within relevant subgroups defined by clinical, social, and functional risk factors that may bias the outcome. Measures that include social or functional risk factors in the final risk model should be calibrated in subgroups defined by those factors to the extent possible. At a minimum, developers should conduct subgroup calibration analyses for race, ethnicity, an indicator of urbanicity/rurality, an indicator of poverty, and an indicator of disability. Developers should be transparent about their approach and their interpretation of the results.

Measure developers can look to **Appendix D** for an example of verbatim text from an NQF-endorsed measure submission. Of note, all four of these NQF-endorsed measures were submitted to NQF prior to the release of this Technical Guidance.

- **Example 4** shows how an NQF-endorsed measure conducted a decomposition analysis to estimate separately the effects on the outcome of the following: (1) patient social risk factors and (2) the proportion of patients within the accountable entity (i.e., hospitals) who have the social risk factor. In other words, this example statistically parses the extent to which the social risk variable reflects an individual-level effect and an accountable entity-level effect. The latter is hypothesized to signal a potential difference in quality of care by social risk factor that could be in the accountable entity's locus of control. However, these types of tests must be taken in context of all other steps in the Technical Guidance, and in and of themselves, they cannot determine whether a variable should or should not be included in the risk model/if risk adjustment is or is not appropriate. The verbatim text in Example 5 shares the developer's examination of its ability to make this determination, given the testing results.
- Additionally, the figure in **Example 5** displays the correlation between a measure outcome with and without social risk adjustment. The figure shows how an NQF-endorsed measure submission has explained determinations about whether adjustment is or is not appropriate based on the marginal effect of adjusting for social risk.
- **Example 6** includes risk decile plots for the purpose of examining risk model calibration across subgroups. The plots display observed versus predicted probabilities for a range of outcomes across quartiles of high to low risk.
- Lastly, **Example 7** uses similar methods for a measure of resource use. Both are examples of how previously NQF-endorsed measures have examined model performance in a range of groups of varying levels of risk.

Considerations for Determining the Final Risk Adjustment Model

Measure developers should examine each measure on a case-by-case basis to determine the appropriateness for social and/or functional risk adjustment, taking into consideration the relationship

between individual risk factors and the outcome in the context illustrated in the conceptual model. This guidance acknowledges that there may be situations in which social and/or functional risk adjustment is either unnecessary or inappropriate. Important considerations include whether the key processes leading to an outcome are directly under the control of the accountable entity. This should be clearly defined in the conceptual model. However, failure to address risk adjustment in an adequate manner can lead to biased conclusions that may adversely affect decision making in research and policy contexts.[63] Additionally, when performance measures are used for accountability applications, such as public reporting and pay for performance, then developers should assess the potential impact on patient populations with social and/or functional risks and the accountable entities serving them to identify and monitor unintended consequences (see the section titled **Negative Unintended Consequences**) and ensure alignment with program and policy goals. Alternatively, inappropriate adjustment for social risk factors has the potential to perpetuate disparities in care by locking in care disparities in quality and incentive programs and reducing the incentives for providers to address disparities. Hence, balanced and thorough consideration and discussion of the trade-offs in adjusting for social and/or functional risk are critical aspects of this standardized framework.

Some categories of measures should not be risk-adjusted for any risk factors, including social and/or functional risk. For example, process measures are generally not risk-adjusted because the process of care being measured should be provided to all patients included in the measure, regardless of patient characteristics. In cases where a measure developer deems it necessary to risk-adjust a process measure, this guidance, although not directed at process measures, can inform that assessment. Further, not all outcome measures may need to be adjusted. Another example is that measures of safety, namely never events, should generally not be risk-adjusted for any risk factors, including social and/or functional risk factors. This is because these events are largely preventable and indicative of inadequacies in a healthcare setting's safety systems, and rates should not be expected to vary with social risk factors, such as poverty, illiteracy, or limited English proficiency. For these measures, developers should stratify by social and functional risk factors to facilitate reporting by subgroups (see the section titled **Stratification** below).

Stratification

Identifying and reducing disparities in health and healthcare are important national priorities and require detailed analysis and reporting of the performance data by patient subgroups. Stratification refers to the division of a population into distinct, mutually exclusive strata or groups defined by particular patient characteristics represented in data, thus enabling the comparison of measure score performance within the specific subgroups.[7] Stratification of patients by social and functional risk factors can be used to more clearly show the areas in which disparities exist or a need is present to identify differences in performance measure results. Thus, future candidate measure endorsement submissions should include measure specifications for reporting measure scores stratified by at-risk populations. These specifications should include variables selected for stratification, an approach to calculating performance measure results, and the reporting approach.

Furthermore, measure developers should consider how the proposed, unstratified measure specifications would be modified to implement the stratification approach. NQF's current **measure**

evaluation criteria require that each measure specification include a description of the measure logic and full details (including numerator and denominator statements), the level of analysis and care settings, a data dictionary and/or Health Quality Measure Format (HQMF) specifications, the type of measure score and appropriate interpretation, a description and copy of the instrument (if an instrument-based measure), the data sources, and component measures and composite construction (if a composite measure). Stratifying a measure impacts multiple components of the specification, and these impacts on the measure's specifications should be documented. It is important to note that depending upon the specific application, results stratified by social and functional risk subgroups may use unadjusted or clinically risk-adjusted (e.g., age, comorbidities) data but not the same stratification variables that have already been used for social or functional risk adjustment. The intent of the stratification recommendation within this Technical Guidance is for stratification to be conducted regardless of decisions for risk adjustment.

Stratification can be a tool for reporting on within-entity differences in measure results between subgroups of patients. For example, CMS currently reports hospital readmission rates publicly as an overall score for each hospital and also confidentially reports performance results for different subgroups or strata of patients at that hospital and across hospitals as a means of comparing entities.[64]

Stratification of within-entity differences for subgroups can promote health equity by identifying opportunities for appropriate resource allocation for quality improvement and for reducing disparities in care delivery. This approach can also influence patient and consumer choice by allowing patients and consumers with social and/or functional risk to see how well accountable entities provide care to patients like them.

Stratification can also be used to compare measure scores calculated for subgroups of patients across accountable entities. This approach is referred to as between-entity differences. For example, performance on the rates of readmissions for a given subgroup (e.g., dual-eligible beneficiaries) can be compared across entities.

Additionally, stratification can also refer to segmenting accountable entities rather than patients. For example, this approach, also known as "peer grouping," is deployed in the Hospital Readmissions Reduction Program (HRRP) and the CMS Star Ratings program. In this approach, performance scores or penalties (in the case of HRRP) are calculated among a set of providers with similar characteristics (e.g., percent of dual-eligible beneficiaries). While these two additional methods for stratification may be appropriate for specific use cases, they may not be feasible in some programs, and measure developers should include patient-level stratification as part of their submission regardless of the intended program use.

The TEP recommended using a minimum set of variables for subgroup stratification analysis, including race, ethnicity, an indicator of urbanicity/rurality, an indicator of poverty (e.g., dual eligibility status), and an indicator of disability (**Minimum Standard #7**). Beyond this minimum set of variables for subgroup stratification analysis, developers should also consider stratification to distinguish between groups of patients who may have difficulties accessing care (e.g., transportation barriers, geographic

distance to a provider, provider shortage areas, and income), as suggested in the literature, by patients, experts, or other stakeholders, and as reflected in the conceptual model.

As mentioned above, the purpose of risk adjustment in measurement is to create measures that fairly compare quality of care between measured entities. Stratification enhances transparency to highlight known areas where disparities exist, which would be researched as part of the construction of the conceptual model. This minimum set of variables guides developers to focus on patient groups with known, widespread disparities. For this subgroup analysis and for the groups identified in the conceptual model, a measure developer's capabilities may also depend on available data and sample sizes within subgroups. If this is the case, measure developers should describe why the analysis was not feasible. This Technical Guidance recognizes that whether programs ultimately use the stratified or unstratified version of measures may not be under the control of measure developers, but specifying measures with a stratification approach can enable the ability to publicly report those results.

---

7 To maximize the ability to identify healthcare disparities, the final measure specifications should provide a stratification approach for calculating and displaying measure scores by at-risk subgroups. At a minimum, developers should, to the extent feasible given the samples' sizes, use race, ethnicity, an indicator of urbanicity/rurality, an indicator of poverty, and an indicator of disability for subgroup stratification analyses. Beyond this minimum set of factors for subgroup stratification analysis, developers should consider stratification to distinguish between groups of patients who may have difficulties accessing care, for example. This should be informed by literature, patients, experts, or other stakeholders and is reflected in the conceptual model. The distribution of the measure scores and sample sizes for each subgroup across providers should also be presented.

Developers should consider the methodological limitations of stratification. For example, depending on the unit of analysis, measured population, etc., stratification can create reliability problems due to small subgroup numbers. This decreases the reliability of estimates in some strata, which generally should not be used for comparative performance evaluation. Developers may choose to mitigate this constraint by lengthening the data collection period, but this has trade-offs for reporting, quality improvement, and actionability. For instance, pooling data across time may mask temporal changes (either improvement or degradations) in performance,[65] which may be of special interest. With this approach, measure results are also less timely, and therefore, they may be less useful for decisions made by the accountable entity and/or program implementer and less useful for tracking near-term impacts of policy incentives. The developer should accompany the stratified data with appropriate explanations about any limitations, any approaches used to mitigate small numbers, or whether there are minimum subgroup sizes. This information is important for payers or quality improvement professionals to know because they would

need to weigh trade-offs between quality improvement program needs and measure reliability to determine how measure results should be reported.

Negative Unintended Consequences

In deciding whether to risk-adjust, developers should consider the potential for negative unintended consequences and determine whether they outweigh the benefits of the performance measure in facilitating progress toward achieving high quality, efficient healthcare. Any approach to measurement should minimize potential negative unintended consequences, especially for patients with social and functional risks.[28]

The first and most important concern about adjustment for social and/or functional risk factors is that disadvantaged patient populations might receive lower quality care. In other words, differences in observed performance across accountable entities reflect actual differences in the processes of care for disadvantaged versus other patients that would be "adjusted away." The concern here is that adjustment will obscure any meaningful differences in quality or performance—that is, the adjustment will have a strong enough effect that meaningful differences in performance will not be detectable in adjusted performance scores. To address this unintended consequence, stakeholders recommended reporting stratified results, as noted above. This stratification will allow the identification of, and facilitate the reduction of, disparities. If an accountable entity's patient population includes a high proportion of disadvantaged patients, it will need to address the needs of that population in order to improve its overall performance.

Adjusting for social and/or functional risk factors may benefit patients with these risk factors, as unadjusted measures may lead to high financial penalties among accountable entities that provide care for patients with a high proportion of social or functional risk and who are unable to mitigate these risks due to limitations in resources (i.e., lack of social work staff, lack of available community resources). These financial penalties may leave some accountable entities who care for disadvantaged populations with fewer resources for quality improvement activities. Failing to account for the greater difficulty in achieving good outcomes in disadvantaged populations could set up a series of adverse feedback loops that result in a downward spiral of access and quality for those populations. For instance, payments may be shifted away (as financial penalties) from accountable entities serving disadvantaged populations and communities and shifted to those serving more affluent, less vulnerable populations and communities (as financial rewards). As a result, accountable entities may have a strong incentive to avoid serving patients with high social or functional risk, so as to avoid being labeled as a "bad performer." Subsequently, individual consumers choosing among accountable entities (e.g., providers, hospitals, and health plans) whose performance will be publicly reported will tend to avoid units serving disadvantaged patients and communities based on performance scores that may not provide a valid comparative performance assessment.

To address these concerns, stakeholders noted that VBP can be designed to mitigate the unintended consequences of unadjusted measures. Due to the differential in resources that may be required to achieve a measured outcome in an at-risk population, payments to accountable entities may also be adjusted so that those serving more at-risk populations would be rewarded more for achieving the same

level of performance as their peers serving more advantaged populations.[66] For example, CMS recently announced the redesigned ACO model, ACO REACH, which incorporates goals of achieving health equity.[10] In 2023, a new cohort of healthcare providers will begin participation in the ACO REACH model, which aims to incentivize ACOs covering underserved communities by making payment adjustments at the beneficiary level using dual-eligibility status and the Area Deprivation Index[11] based on the census block of residence. As the ACO REACH model matures, there is the potential for replacing measure-level adjustment with adjustment within the payment model design.

Overall, to mitigate any impact of unintended consequences on patients with social and/or functional risk factors, developers should continuously monitor the measure once implemented for potential unintended, adverse consequences. Current **NQF measure evaluation criteria** require performance measurement to facilitate progress toward achieving high quality, efficient healthcare for individuals or populations; this progress should also consider any unintended negative consequences to individuals or populations (if such evidence exists). NQF evaluates this type of information when it reviews measures for re-endorsement. It encourages measure developers and stewards to evaluate and report on any unintended consequences, including the nature of the consequence, the affected party, the number of people affected, and the severity of the impact.

Policy Considerations

Looking forward, there are critical NQF policy and broader healthcare policy considerations that emerge from this Technical Guidance. As it relates to NQF policy, important implementation steps need to be considered prior to the full implementation of the minimum standards and recommendation into NQF's **Consensus Development Process** endorsement process. NQF will need to inform and garner feedback from the broad NQF stakeholder community (i.e., NQF members organizations, the CSAC, the SMP, measure developers, and NQF Standing Committees) on how this Technical Guidance will be further operationalized for both measure development and endorsement.

NQF Policy

This Technical Guidance was developed by working with stakeholders to define a standardized and science-based framework for considering social and/or functional risk factors in the development of risk adjustment models used in healthcare quality measurement. This approach sets forth a five-step process for developing risk models and defines seven risk adjustment standards. While this project gathered input through an environmental scan and from a multistakeholder TEP, focus groups, and public comment, specific changes to NQF's CDP[67] require several important steps prior to the implementation of this guidance, including the following:

- Translating the five-step process outlined in the Technical Guidance and the associated risk adjustment standards into standard NQF measure information specifications collected for candidate measures considered for measure endorsement
- Updating the NQF measure endorsement criteria to reflect the risk adjustment standards outlined in the Technical Guidance

- Coordinating with CMS to implement potential updates to the CMS Quality Measures Blueprint[68] and with other measure developers to ensure the alignment of measure development guidelines and endorsement standards
- Educating and gaining further feedback from stakeholders, specifically measure developers, on proposed updates to the NQF measure submission documents and updates to the NQF measure endorsement criteria

Broadly, there are several areas of NQF's measure information specifications and the validity component of the NQF Scientific Acceptability of Measure Priorities endorsement criterion that will require updating. These will potentially include the following:

- An explanation of the conceptual model used to design the risk adjustment model
- An examination of specific, critical social and/or functional risk factors that should be considered in the conceptual model
- If possible, descriptive statistics on how the risk variables identified from the conceptual model are distributed across the measured entities
- Model calibration conducted within the overall population and within relevant at-risk clinical, social, and functional subgroups for all risk-adjusted measures
- Specifications on how the measures should be stratified, regardless of whether a social or functional risk factor is included in the final risk adjustment model

Furthermore, the current NQF endorsement criteria are agnostic to measure use. This 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 current purview of NQF endorsement. Implementing the approach to risk adjustment model review suggested by this report would require NQF to modify its criteria.

NQF acknowledges that approaches to addressing social and functional risk in risk adjustment models will be an area that continues to evolve with new data sources, advancements in measurement science, and continued monitoring and feedback on measures in use. Thus, NQF will consider how to incorporate continued monitoring and feedback into the NQF measure maintenance process.

Healthcare Policy

Quality measures are first and foremost a tool to improve care for patients by drawing inferences on accountable-entity performance on the measured outcome. A key focus of improvement should be reducing disparities in care. This Technical Guidance seeks to advance the science and policy of whether and how to adjust quality measures for social and functional risk factors so that accountable entities will be compared fairly. It guides the development of risk adjustment outcome and cost/resource use measures that fairly evaluate the accountable entities that serve at-risk populations. It provides an important but narrow component of the measurement strategy we need. Although this guidance does not inform how measures are implemented, it is important to signal that the impact of program polices on accountable entities caring for disadvantaged populations should be considered. These accountable entities may have fewer resources to improve the care they provide (i.e., few community resources,

social work staff). Therefore, a broader healthcare policy must have a holistic approach to addressing disparities. The recent ACO REACH model is one example of steps some payers are taking toward this holistic approach by implementing risk-adjusted payment approaches that aim to incentivize accountable entities to better support care delivery and coordination for people in underserved communities.[10]

To support these new care delivery models and policies, measurement organizations can identify individual performance measures that can serve to identify disparities in care using disparities-sensitive measures. 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."[31]

In addition to identifying current performance measures that can be classified as "disparities-sensitive," performance measures need to be developed to directly measure health equity.[69] The RWJF lays out the need for short-, intermediate-, and long-term measures of health equity that should be pursued. Direct measures of social exclusion, marginalization, discrimination, and disadvantage[69] are critical examples of measurement areas in need of further work. Efforts to advance health equity require a holistic measurement approach, which combines refinement of approaches to risk adjustment for social and functional risk, identification of disparities-sensitive measures, and the development of health equity measures.

Lastly, beyond individual performance measures, accountability programs should consider whether and how to incorporate these social and/or functional risk factors into comparative evaluations for the purposes of assigning rewards and penalties.[70] These accountable entities may have fewer resources to improve the care they provide. Quality improvement programs can provide support to accountable entities in other ways. This could include additional payments to safety-net providers and bonuses to those who demonstrate high quality care for patients with higher social and/or functional risk.[27,71] Although they are used for different purposes, there are already existing payments and bonuses that target safety-net providers, including the current payments and bonus points for small practices and practices with a higher share of medically and socially complex patients in the MIPS program.[72]

Conclusion

As the U.S. continues to move toward value-based care, the need is correspondingly greater to advance the field of measurement science and ensure that performance measurement is unbiased and accurate. The increased use of outcome and cost/resource use measures in payment models and public reporting programs has resulted in greater scrutiny regarding the adequacy and fairness of the risk adjustment methodologies for measured accountable entities, especially as it relates to social and functional risk factors. Risk-adjusting outcome performance measures (inclusive of cost/resource use) to account for differences in patient health that affect outcomes is widely accepted. Additionally, with social and functional risk factor adjustment being absent from certain performance measures, accountable entities

may avoid caring for the most at-risk and disadvantaged patients because of their anticipated worse outcomes or higher costs, potentially worsening inequities. However, concerns exist that adjusting for social risk may excuse delivery of lower-quality care to socially at-risk populations and that lower performance is masked with statistical adjustment. These differing perspectives, along with variation in data sources and risk adjustment methods and approaches for similar measures, have led to an increased need for standardization.

Building on several years of guidance for risk adjustment model development, NQF sought to advance measurement science by developing technical guidance for measure developers, which includes agreed-upon best practices on when and how to adjust for functional and social risk factors in measure development. These best practices, referred to within this guidance as a set of standards that developers should follow, at a minimum, apply to both outcome and cost/resource use performance measures at any level of analysis (e.g., health plans, facilities, individual clinicians, and ACOs).

During the initial phase of the project, NQF conducted an environmental scan to identify current uses of functional and social risk factors in measurement. Findings were used to inform the decision making of an NQF-convened, multistakeholder TEP, which identified seven sequential minimum standards rooted in core principles of quality measurement and risk adjustment science. During the second phase of this project, NQF broadened stakeholder engagement efforts to garner input on the utility of the Technical Guidance and to potentially expand the discussion to topics on which stakeholders have sought clarity or more detail. Stakeholder groups included patient and consumers, providers, public- and private-sector payers/purchasers, measure developers, and NQF measure evaluation bodies (e.g., the SMP, the CSAC, and Standing Committees). Additionally, NQF engaged with individuals from historically underserved communities and those with minority viewpoints (i.e., those who had expressed comments disagreeing with the Technical Guidance recommendations and/or minimum standards) to elicit their input and rationale for the TEP's consideration. This expanded stakeholder engagement resulted in an increased focus and attention on the role of quality measurement and risk adjustment approaches in promoting health equity and reducing disparities.

This step-by-step guidance for social and/or functional risk factor adjustment includes the evaluation of a conceptual and empirical relationship to the outcome being measured. The initial phase TEP emphasized the importance of first establishing a robust conceptual model that considers a minimum set of social and functional risk factors. The guidance for selecting risk factors for adjustment, along with statistical and epidemiological theory and practices, provides a prudent basis for making determinations for social and/or functional risk adjustment. Furthermore, to mitigate concerns that risk adjustment masks disparities in care, this guidance instructs developers to stratify measure results by key risk variables to identify healthcare disparities and further promote health equity.

Risk adjustment is not perfect; the same limitation that occurs when adjusting for clinical factors applies to social and functional risk factors (i.e., risk adjustment can only account for measurable and reportable factors). Additionally, risk adjustment procedures only address patient characteristics, and there could be accountable-entity characteristics (e.g., funding of safety-net providers, area healthcare workforce, and community resources) that might have policy implications related to some accountability

applications. However, this guidance will facilitate consistency in the evaluation of these risk models through a set of standard best practices that promote transparency and innovation within measurement science. Furthermore, this work may have implications for the review and consideration of measures for use within public reporting and accountability applications. Namely, as social and functional risks are increasingly recognized as having a tremendous impact on health and healthcare outcomes, measurement organizations can identify individual performance measures that can help to identify disparities in care using disparities-sensitive measures. Additionally, the inclusion of these factors in risk adjustment model considerations may further promote health equity and transparency of health disparities. However, this requires a systematic, holistic policy approach to fully addressing inequities associated with race, ethnicity, and other social risks. As a healthcare quality standard-setting entity, NQF continues in its efforts to promote health equity and advance measurement science.

References

1 Centers for Medicare & Medicaid Services. What is the CMS National Quality Strategy? | CMS. What is the CMS National Quality Strategy? https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/CMS-Quality-Strategy. Published June 10, 2022. Last accessed June 2022.

2 Johnston KJ, Joynt Maddox KE. The Role Of Social, Cognitive, And Functional Risk Factors In Medicare Spending For Dual And Nondual Enrollees. *Health Aff (Millwood)*. 2019;38(4):569-576. https://www.healthaffairs.org/doi/10.1377/hlthaff.2018.05032. Last accessed November 2022.

3 Michelle Eder, PhD, Michelle Henninger, PhD, Shauna Durbin, MPH, et al. Screening and Interventions for Social Risk Factors: Technical Brief to Support the US Preventive Services Task Force | Guidelines | JAMA | JAMA Network https://jamanetwork.com/journals/jama/fullarticle/2783975. Published September 1, 2021. Last accessed June 2022.

4 Kreuter MW, Thompson T, McQueen A, et al. Addressing Social Needs in Health Care Settings: Evidence, Challenges, and Opportunities for Public Health. *Annu Rev Public Health*. 2021;42:329-344.

5 National Association of ACOs (NAACOS). The Role of ACOs in Addressing Health Equity. https://www.naacos.com/assets/docs/pdf/2021/ACOsandHealthEquityWhitePaper092121.pdf. Published nd.

6 Michael E. Chernew, J. Michael McWilliams. The Case For ACOs: Why Payment Reform Remains Necessary | Health Affairs Forefront. https://www.healthaffairs.org/do/10.1377/forefront.20220120.825396/full/. Published January 24, 2022. Last accessed June 2022.

7 National Quality Forum (NQF). NQF: Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors. https://www.qualityforum.org/Publications/2014/08/Risk_Adjustment_for_Socioeconomic_Status_or_Other_Sociodemographic_Factors.aspx. Published August 2014. Last accessed August 2022.

8 Centers for Medicare & Medicaid Services. Hospital Readmissions Reduction Program (HRRP) Methodology. https://qualitynet.cms.gov/inpatient/hrrp/methodology. Last accessed December 2022.

9 Centers for Medicare & Medicaid Services. Federal Register :: Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long-Term Care Hospital Prospective Payment System and Policy Changes and Fiscal Year 2018 Rates; Quality Reporting Requirements for Specific Providers; Medicare and Medicaid Electronic Health Record (EHR) Incentive Program Requirements for Eligible Hospitals, Critical Access Hospitals, and Eligible Professionals; Provider-Based Status of Indian Health Service and Tribal Facilities and Organizations; Costs Reporting and Provider Requirements; Agreement Termination Notices. https://www.federalregister.gov/documents/2017/08/14/2017-16434/medicare-program-hospital-inpatient-prospective-payment-systems-for-acute-care-hospitals-and-the#p-3010. Last accessed November 2022.

10 Centers for Medicare & Medicaid Services. CMS Redesigns Accountable Care Organization Model to Provide Better Care for People with Traditional Medicare. CMS Redesigns Accountable Care Organization Model to Provide Better Care for People with Traditional Medicare. https://www.cms.gov/newsroom/press-releases/cms-redesigns-accountable-care-organization-model-provide-better-care-people-traditional-medicare. Published February 24, 2022.

11 Neighborhood Atlas. Neighborhood Atlas - Mapping. https://www.neighborhoodatlas.medicine.wisc.edu/mapping. Published 2022. Last accessed June 2022.

12 Sahil Sandhu, Robert S. Saunders, Mark B. McClellan, et al. Health Equity Should Be A Key Value In Value-Based Payment And Delivery Reform | Health Affairs Forefront. https://www.healthaffairs.org/do/10.1377/forefront.20201119.836369/full/. Published November 25, 2020. Last accessed June 2022.

13 Shahian DM, Badhwar V, O'Brien SM, et al. Social Risk Factors in Society of Thoracic Surgeons Risk Models. Part 1: Concepts, Indicator Variables, and Controversies. *Ann Thorac Surg*. 2022;113(5):1703-1717. https://www.scopus.com/inward/record.url?scp=85125476215&partnerID=8YFLogxK. Last accessed August 2022.

14 Iezzoni LI, ed. *Risk Adjustment for Measuring Health Care Outcomes*. 4th ed. Chicago, IL: Health Administration Press; 2013.

15 Franks P, Fiscella K. Effect of Patient Socioeconomic Status on Physician Profiles for Prevention, Disease Management, and Diagnostic Testing Costs. *Med Care*. 2002;40:717-724.

16 Blum A, Egorova N, Sosunov EA. Impact of Socioeconomic Status Measures on Hospital Profiling in New York City. *Circ Cardiovasc Qual Outcomes*. 2014;7:391-397.

17 Bierman AS, Lawrence WF, Haffer SC, et al. Functional health outcomes as a measure of health care quality for Medicare beneficiaries. *Health Serv Res*. 2001;36(6 Pt 2):90-109.

18 Institute of Medicine (US) Committee on Quality of Health Care in America. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington (DC): National Academies Press (US); 2001. Accessed February 2021 at http://www.ncbi.nlm.nih.gov/books/NBK222274/.

19 Deutsch A, Kline T, Kelleher C, et al. Analysis of Crosscutting Medicare Functional Status Quality Metrics Using the Continuity and Assessment Record and Evaluation (CARE) Item Set. Accessed February 2021 at https://aspe.hhs.gov/basic-report/analysis-crosscutting-medicare-functional-status-quality-metrics-using-continuity-and-assessment-record-and-evaluation-care-item-set. Published June 13, 2015. Last accessed February 2021.

20 Bernheim SM, Parzynski CS, Horwitz L. Accounting for Patients' Socioeconomic Status Does not Change Hospital Readmission Rates. *Health Aff (Millwood)*. 2016;35:1461-1470.

21 Chatterjee P, Werner RM. The Hospital Readmission Reduction Program and Social Risk. *Health Serv Res*. 2019;54:324-326.

22 Tirupathi R, Muradova V, Shekhar R, et al. COVID-19 disparity among racial and ethnic minorities in the US: A cross sectional analysis. *Travel Med Infect Dis*. 2020;38:101904.

23 Wang Q, Berger NA, Xu R. Analyses of Risk, Racial Disparity, and Outcomes Among US Patients With Cancer and COVID-19 Infection. *JAMA Oncol*. 2021;7(2):220-227.

24 Wang Q, Davis P, Gurney M, et al. COVID-19 and dementia: Analyses of risk, disparity, and outcomes from electronic health records in the US. *Alzheimers Dement J Alzheimers Assoc*. February 2021.

25 NQF. Social Risk Trial Final Report. https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=95208. Last accessed May 2021.

26 National Quality Forum (NQF). NQF: A Strategic Plan for Achieving The Care We Need. https://www.qualityforum.org/A_Strategic_Plan_for_Achieving_The_Care_We_Need.aspx. Last accessed June 2022.

27 ASPE. Second Report to Congress on Social Risk and Medicare's Value-Based Purchasing Programs. ASPE. Accessed at https://aspe.hhs.gov/pdf-report/second-impact-report-to-congress. Published June 29, 2020. Last accessed June 2021.

28 National Academies of Sciences, Engineering, and Medicine. *Accounting for Social Risk Factors in Medicare Payment: Identifying Social Risk Factors*. Washington, D.C.: National Academies Press; 2016. http://www.nap.edu/catalog/21858. Last accessed January 2021.

29 National Quality Forum (NQF). NQF: Adjusting Measures for Social Risk Factors. Accessed January 2021 at https://www.qualityforum.org/Adjusting_Measures_for_Social_Risk_Factors.aspx.

30 National Quality Forum. Evaluation of the NQF Trial Period for Risk Adjustment for Social Risk Factors. *July 2017*. Accessed at https://www.qualityforum.org/Publications/2017/07/Social_Risk_Trial_Final_Report.aspx.

31 NQF. NQF: Healthcare Disparities and Cultural Competency Consensus Standards: Disparities-Sensitive Measure Assessment. https://www.qualityforum.org/Publications/2012/11/Healthcare_Disparities_and_Cultural_Competency_Consensus_Standards__Disparities-Sensitive_Measure_Assessment.aspx. Published 2012. Last accessed May 2022.

32 World Health Organization, ed. *International Classification of Functioning, Disability and Health: ICF*. Geneva: World Health Organization; 2001. https://psychiatr.ru/download/1313?view=name=CF_18.pdf.

33 National Committee on Vital and Health Statistics (NCVHS), Subcommittee on Populations. *Classifying and Reporting Functional Status*. Accessed at https://www.ncvhs.hhs.gov/wp-content/uploads/2017/08/010617rp.pdf.

34 Kaiser Family Foundation. Disparities in Health and Health Care: 5 Key Questions and Answers. Disparities in Health and Health Care. Accessed June 2021 at

https://www.kff.org/racial-equity-and-health-policy/issue-brief/disparities-in-health-and-health-care-5key-question-and-answers/. Published May 11, 2021. Last accessed June 2021.

35 Culyer A. Equity - some theory and its policy implications. *J Med Ethics*. 2001;27(4):275-283.

36 Libert Y, Janne P, Razavi D, et al. Impact of medical specialists' locus of control on communication skills in oncological interviews. *Br J Cancer*. 2003;88(4):502-509. https://www.nature.com/articles/6600797. Last accessed June 2022.

37 Centers for Medicare & Medicaid Services. CMS Quality Measurement Action Plan: March 2021. CMS Quality Measurement Action Plan. https://www.cms.gov/files/document/2021-cms-quality-conference-cms-quality-measurement-action-plan-march-2021.pdf. Published March 2021.

38 National Quality Forum (NQF). NQF: Risk Adjustment Guidance Final Environmental Scan. Accessed May 2021 at https://www.qualityforum.org/Publications/2021/05/Risk_Adjustment_Guidance_Final_Environmental_Scan.aspx. Last accessed May 2021.

39 Boulkedid R, Abdoul H, Loustau M, et al. Using and Reporting the Delphi Method for Selecting Healthcare Quality Indicators: A Systematic Review. *PLoS ONE*. 2011;6(6). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3111406/. Last accessed April 2021.

40 McLaughlin JA, Jordan GB. Using Logic Models. In: *Handbook of Practical Program Evaluation*. Fourth. John Wiley & Sons, Ltd; 2015:62-87. Accessed April 2021 at https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119171386.ch3.

41 Zook M, Green K. When Talking About Social Determinants, Precision Matters. Health Affairs Blog. Accessed April 2021 at https://www.healthaffairs.org/do/10.1377/hblog20191025.776011/full/.

42 Garfield R, Tolbert J, 2021. Medicaid Enrollment Churn and Implications for Continuous Coverage Policies. *KFF*. December 2021. https://www.kff.org/medicaid/issue-brief/medicaid-enrollment-churn-and-implications-for-continuous-coverage-policies/. Last accessed August 2022.

43 The Gravity Project. HL7 International. Accessed at https://www.hl7.org/gravity/. Published April 22, 2021. Last accessed May 2021.

44 The PACIO Project. The PACIO Project. Accessed at http://pacioproject.org/. Published April 22, 2021. Last accessed May 2021.

45 Guo Y, Chen Z, Xu K, et al. International Classification of Diseases, Tenth Revision, Clinical Modification social determinants of health codes are poorly used in electronic health records. *Medicine (Baltimore)*. 2020;99(52):e23818. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7769291/. Last accessed June 2022.

46 National Association of Community Health Centers. Protocol for Responding to and Assessing Patients' Assets, Risks, and Experiences (PRAPARE). NACHC. Accessed May 2021 at https://www.nachc.org/research-and-data/prapare/. Published April 22, 2021.

47 US Census Bureau. US Census Bureau. Census.gov. Accessed May 2021 at https://www.census.gov/en.html. Last accessed June 2021.

48 US Census Bureau. American Community Survey (ACS). The United States Census Bureau. Accessed May 2021 at https://www.census.gov/programs-surveys/acs. Last accessed June 2021.

49 University of Wisconsin School of Medicine and Public Health. Making Neighborhood Disadvantage Metrics Accessible: The Neighborhood Atlas. The Neighborhood Atlas. Accessed May 2021 at https://www.neighborhoodatlas.medicine.wisc.edu/.

50 Fremont A, Weissman J, Hoch E, et al. When Race/Ethnicity Data Are Lacking. When Race/Ethnicity Data Are Lacking, Using Advanced Indirect Estimation Methods to Measure Disparities. Accessed at https://www.rand.org/pubs/periodicals/health-quarterly/issues/v6/n1/16.html. Last accessed June 2021.

51 Sandelowski M. Whatever happened to qualitative description? *Res Nurs Health*. 2000;23(4):334-340.

52 Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the Performance of Prediction Models: A Framework for Traditional and Novel Measures. *Epidemiology*. 2010;21(1):128-138. Accessed May 2021 at https://journals.lww.com/00001648-201001000-00022. Last accessed May 2021.

53 Pencina MJ, D'Agostino RB, D'Agostino RB, et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008;27(2):157-172; discussion 207-212.

54 Deutscher D, Werneke M, Hayes D, et al. Impact of Risk Adjustment on Provider Ranking for Patients With Low Back Pain Receiving Physical Therapy. *J Orthop Sports Phys Ther*. 2018;48(8):637-648.

55 Mihaescu R, van Zitteren M, van Hoek M, et al. Improvement of risk prediction by genomic profiling: reclassification measures versus the area under the receiver operating characteristic curve. *Am J Epidemiol*. 2010;172(3):353-361.

56 Mühlenbruch K, Heraclides A, Steyerberg EW, et al. Assessing improvement in disease prediction using net reclassification improvement: impact of risk cut-offs and number of risk categories. *Eur J Epidemiol*. 2013;28(1):25-33.

57 Leening MJG, Vedder MM, Witteman JCM, et al. Net reclassification improvement: computation, interpretation, and controversies: a literature review and clinician's guide. *Ann Intern Med*. 2014;160(2):122-131.

58 Pepe MS, Fan J, Feng Z, et al. The Net Reclassification Index (NRI): a Misleading Measure of Prediction Improvement Even with Independent Test Data Sets. *Stat Biosci*. 2015;7(2):282-295.

59 Cook NR, Paynter NP. Performance of reclassification statistics in comparing risk prediction models. *Biom J Biom Z*. 2011;53(2):237-258.

60 Hilden J, Gerds TA. A note on the evaluation of novel biomarkers: do not rely on integrated discrimination improvement and net reclassification index. *Stat Med*. 2014;33(19):3405-3414.

61 David W. Hosmer Jr., Stanley Lemeshow, Rodney X. Sturdivant. *Applied Logistic Regression, 3rd Edition | Wiley*.; 2013. https://www.wiley.com/en-us/Applied+Logistic+Regression%2C+3rd+Edition-p-9780470582473. Last accessed August 2022.

62 Centers for Medicare & Medicaid Services. ACO #8 – Risk Standardized All Condition Readmissions. https://www.cms.gov/medicare/medicare-fee-for-service-payment/sharedsavingsprogram/downloads/aco-8.pdf. Published 2016.

63 Vogel W, Chen GJ. An introduction to the why and how of risk adjustment. *Biostat Epidemiol*. 2020;4(1):84-97. Accessed May 2021 at https://doi.org/10.1080/24709360.2018.1519990.

64 Centers for Medicare & Medicaid Services. Hospital Readmissions Reduction Program (HRRP) | CMS. https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program. Published August 5, 2022. Last accessed August 2022.

65 Joynt Maddox K, Reidhead M, Qi A, et al. Association of Stratification by Dual Enrollment Status With Financial Penalties in the Hospital Readmissions Reduction Program. *JAMA Intern Med*. 2019;179(6):769.

66 Addressing Social Risk Factors In Value-Based Payment: Adjusting Payment Not Performance To Optimize Outcomes and Fairness | Health Affairs. https://www.healthaffairs.org/do/10.1377/hblog20210414.379479/full/. Last accessed August 2021.

67 National Quality Forum (NQF). NQF: Consensus Development Process. https://www.qualityforum.org/measuring_performance/consensus_development_process.aspx. Last accessed June 2022.

68 Center for Medicare & Medicaid Services. *CMS Measures Management System Blueprint Supplemental Material: Risk Adjustment in Quality Measurement*.; 2020. Accessed March 2021 at https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjNqLm-sKbwAhXHKFkFHbOXCCUQFjABegQIBBAD&url=https%3A%2F%2Fwww.cms.gov%2Ffiles%2Fdocument%2Fblueprint-risk-adjustment.pdf&usg=AOvVaw0lc_0K-aZNbUARysY4tUWD.

69 Braveman P, Arkin E, Orleans T, et al. What is Health Equity? RWJF. https://www.rwjf.org/en/library/research/2017/05/what-is-health-equity-.html. Published May 1, 2017. Last accessed May 2022.

70 National Academies of Sciences, Engineering, and Medicine. *Accounting for Social Risk Factors in Medicare Payment*. Washington, D.C.: National Academies Press; 2017. https://www.nap.edu/catalog/23635/accounting-for-social-risk-factors-in-medicare-payment. Last accessed August 2021.

71 Fremont A, Weissman J, Hoch E, et al. When Race/Ethnicity Data Are Lacking. When Race/Ethnicity Data Are Lacking, Using Advanced Indirect Estimation Methods to Measure Disparities. Accessed at https://www.rand.org/pubs/periodicals/health-quarterly/issues/v6/n1/16.html. Last accessed June 2021.

72 National Quality Forum (NQF). NQF: NQF Issues Quality Roadmap for Reducing Healthcare Disparities. https://www.qualityforum.org/NQFs_Roadmap_to_Health_Equity.aspx. Published 2017. Last accessed May 2022.

73 DeCarlo M, Cummings C, Agnelli K. Bivariate analysis. In: *Graduate Research Methods in Social Work*. Open Social Work; 2020. Accessed May 2021 at https://viva.pressbooks.pub/mswresearch/chapter/15-bivariate-analysis/.

74 Marcello Pagano, Kimberlee Gauvreau. *Principles of Biostatistics*. Second.; 2000.

75 Office for Civil Rights (OCR). A Decision Tool: Data Use Agreement. HHS.gov. Accessed May 2021 at https://www.hhs.gov/hipaa/for-professionals/special-topics/emergency-preparedness/data-use-agreement/index.html. Published August 29, 2007.

76 World Health Organization. International Classification of Functioning, Disability and Health (ICF). https://www.who.int/classifications/international-classification-of-functioning-disability-and-health.

77 Research 101: Generalizability. Accessed May 2021 at https://www.hydroassoc.org/research-101-generalizability/. Last accessed May 2021.

78 Fast Healthcare Interoperability Resources® (FHIR) | eCQI Resource Center. Accessed May 2021 at https://ecqi.healthit.gov/fhir.

79 Kraemer H, Stice E, Kazdin A, et al. How Do Risk Factors Work Together? Mediators, Moderators, and Independent, Overlapping, and Proxy Risk Factors. *Am J Psychiatry*. 2001;158(6):848-856.

80 Nerenz D, Austin J, Deutscher D, et al. Adjusting Quality Measures For Social Risk Factors Can Promote Equity In Health Care. *Health Aff (Millwood)*. 2021;40(4):637-644.

81 Centers for Medicare & Medicaid Services. *Trends in Part C & D Star Rating Measure Cut Points*.; 2020. Accessed at https://www.cms.gov/files/document/2021-cut-point-trend.pdf.

Appendix A: TEP Members, Federal Liaisons, CMS, and NQF Staff

TEP Members

Philip Alberti, PhD *(Co-Chair)*
Senior Director, Health Equity Research and Quality, Association of American Medical Colleges
Washington, District of Columbia

Karen Joynt Maddox, MD, MPH *(Co-Chair)*
Assistant Professor, Washington University School of Medicine
St. Louis, Missouri

Arlene Ash, PhD
Professor and Division Chief, Biostatistics and Health Services Research, Department of Population and Quantitative Health Sciences, University of Massachusetts Medical School
Boston, Massachusetts

Susannah Bernheim, MD, MHS
Associate Professor; Director, Quality Measurement Program (CORE); Assistant Clinical Professor Section of General Internal Medicine; Core Faculty, Robert Wood Johnson Clinical Scholars Program
New Haven, Connecticut

Melissa Castora-Binkley, PhD, MA, PMP
Pharmacy Quality Alliance (PQA)
Alexandria, Virginia

Lukejohn Day, MD
Chief Medical Officer, Zuckerberg San Francisco General Hospital (ZSFG); Professor of Medicine, University of California San Francisco (UCSF)
San Francisco, California

Marc Elliott, PhD, MA
Senior Principal Researcher, The RAND Corporation
Santa Monica, California

Rachel Harrington, PhD
Research Scientist, National Committee for Quality Assurance
Washington, District of Columbia

Vincent Liu, MD, MS
Research Scientist, Kaiser Permanente Division of Research
Oakland, California

Danielle Lloyd, MPH
SVP Private Market Innovations & Quality Initiatives, America's Health Insurance Plans

Washington, District of Columbia

John Martin, PhD, MPH
Vice President, Data Science, Premier Healthcare Alliance
Charlotte, North Carolina

Shalini Prakash, MS
Data Scientist, Washington Healthcare Authority
Olympia, Washington

Sandra Richardson, MS
Research Scientist; Director, Bureau of Quality Measurement of Special Populations
Albany, New York

Clarke Ross, DPA
Public Policy Director, American Association on Health and Disability (AAHD)
Rockville, Maryland

David Shahian, MD
Vice President, Center for Quality and Safety, Massachusetts General Hospital
Boston, Massachusetts

Cristie Upshaw Travis, MSHHA
Chief Executive Officer, Memphis Business Group on Health (MBGH)
Memphis, Tennessee

Janice Tufte
Hassanah Consulting
Seattle, Washington

Federal Liaisons

Shafa Al-Showk, PhD
Center for Medicare/Centers for Medicare & Medicaid Services (CM/CMS)

Joel Andress, PhD
Center for Clinical Standards and Quality/Centers for Medicare & Medicaid Services (CCSQ/CMS)

Craig Caplan, MA
Federal Office of Rural Health Policy
Health Resources and Services Administration (HRSA)

Maushaumi (Mia) DeSoto, PhD
Center for Quality Improvement and Patient Safety
Agency for Healthcare Research and Quality (AHRQ)

Andy Frankos-Rey, MA
Center for Medicaid & CHIP Services/Centers for Medicare & Medicaid Services (CMCS/CMS)

Sarah Gaillot, PhD
Center for Medicare/Centers for Medicare & Medicaid Services (CM/CMS)

David Nyweide, PhD
Center for Medicare & Medicaid Innovation/Centers for Medicare & Medicaid Services (CMMI/CMS)

Jesse Roach, MD
Center for Clinical Standards and Quality/Centers for Medicare & Medicaid Services (CCSQ/CMS)

Lok Wong Samson, PhD
Office of Health Policy
Office of the Assistant Secretary for Planning and Evaluation (ASPE)

Rachael Zuckerman, PhD
Office of Health Policy
Office of the Assistant Secretary for Planning and Evaluation (ASPE)

CMS

Laura de Nobel, JD, RN
TO COR, DPMS, Centers for Medicare & Medicaid Services

Sophia Chan, PhD, MPH
DVIQR, Center for Medicare/Centers for Medicare & Medicaid Services (CM/CMS)

Helen Dollar-Maples, RN, MSN
Director, DPMS, Centers for Medicare & Medicaid Services

Nidhi Singh-Shah, MPH
Deputy Director, DPMS, Centers for Medicare & Medicaid Services

Gequincia Polk, MPA
IDIQ COR, DPMS, Centers for Medicare & Medicaid Services

Marsha Smith, MD, MPH
Medical Officer, DPMS, Centers for Medicare & Medicaid Services

NQF Staff

Elizabeth Drye, MD, SM
Chief Scientific Officer, Measurement Science and Application

Tricia Elliott, DHA, MBA, CPHQ, FNAHQ
Senior Managing Director, Measurement Science and Application

Matthew K. Pickering, PharmD
Senior Director, Measurement Science and Application

Taroon Amin, PhD, MPH
Consultant

Sai Ma, PhD
Consultant

Monika Harvey, MBA, PMP
Senior Project Manager, Emerging Initiatives and Program Operations

Hannah Ingber, MPH
Manager, Measurement Science and Application

Simone Bernateau
Analyst, Emerging Initiatives and Program Operations

Tristian Wind, BS
Analyst, Measurement Science and Application

Appendix B: Glossary

Accountable entity refers to an individual health professional, health facility, health plan, or health organization/facility that is responsible or accountable for healthcare quality, outcomes, or cost of care.

Area-level variables are those whose unit of measurement/observation is attributed to a geographic unit/level. For example, country, state, county, ZIP code (+4), etc.

Bivariate analyses consist of a group of statistical techniques that examine the relationship between two variables.[73]

Between-unit differences occur when measured entities have different case mix, and quality varies between these measured entities (e.g., a hospital providing lower quality care for a large number of socially disadvantaged patients compared with a hospital with fewer disadvantaged patients exhibiting between-unit differences).

Collinearity refers to the relationship between two variables when one is highly linearly correlated with the other.[74]

Confounders refer to variables that are related to both the intervention and the measured outcome.[14]

Data Use Agreement (DUA) establishes who is permitted to use and receive the various types of data files and the permitted uses and disclosures of such information by the recipient, provided that the recipient will not use or disclose the information other than as permitted by the DUA or as otherwise required by law. A DUA further establishes appropriate safeguards to prevent uses or disclosures of the information that are inconsistent with the DUA and ensures that any agents to whom it provides the limited data sets (LDSs) agree to the same restrictions and conditions that apply to the LDS recipient.[75]

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 quality among populations or social groupings (race, ethnicity, language, etc.).[31]

Endogenous variable refers to a factor in a model whose value is determined by the states of other variables in the model.

Exogenous variable refers to a factor in a model whose value is not determined by the states of other variables in the model.

Functional status is variously defined in the health field. Generally, functional status refers to an attribute that assesses how a health condition has had an impact on an individual's body function, body structures, and ability to participate in activities and complete basic daily tasks.[76] Functional status covers both the individual carrying out ADLs and the individual participating in life situations and society.[33] This includes basic physical and cognitive activities, such as walking or reaching, focusing attention, and communicating, as well as the routine ADLs, including eating, bathing, dressing, transferring, and toileting. This also includes life situations, such as school or play for children, and for

adults, working outside the home or maintaining a household. Furthermore, functional limitations occur when a person's capacity to carry out such activities or performance of such activities is compromised due to a health condition or injury and is not compensated by environmental factors (including physical, social, and attitudinal factors). Functional status encompasses the whole person and is affected by physical, developmental, behavioral, emotional, social, and environmental conditions.[32]

Functional status-related risk adjustment refers to statistical adjustment for functional status-related variables (e.g., frailty, disability, ADLs, cognitive function).

Generalizability is a measure of how useful the results of a study are for a broader group of people or situations. If the results of a study are broadly applicable to many different types of people or situations, the study is said to have good *generalizability*.[77]

Healthcare disparities refer to differences between groups in health insurance coverage, access to and use of care, and quality of healthcare services.[34]

Health disparities refer to a higher burden of illness, injury, disability, or mortality experienced by one group relative to another.[34]

Health equity is the principle underlying a commitment to reduce—and ultimately eliminate—disparities in health and in its determinants, including social determinants. Health equity strives to ensure everyone has a fair and just opportunity to be as healthy as possible. This requires removing obstacles to health, such as poverty, discrimination, and their consequences, including powerlessness and lack of access to good jobs with fair pay, quality education and housing, safe environments, and health.[35]

Health Level Seven (HL7) Fast Health Interoperability Resource (FHIR) refers to the HL7 International standard for exchanging healthcare information electronically. FHIR provides a means for representing and sharing information among clinicians and organizations in a standard way, regardless of the ways local EHRs represent or store the data.[78]

Locus of control refers to the scope of actions that the accountable entity can take to influence the measured outcome.[36] For example, clinicians caring for patients in small rural practices may have a limited ability to address and reduce the impact of social risk factors, such as transportation barriers and access to healthy foods, whereas larger, more urban providers, such as accountable care organizations, may be both incentivized and resourced to address these needs.

Mediator variable refers to a variable within the causal pathway between the actions of the accountable entity and the measured outcome. In this context, an accountable entity action influences the mediator, which in turn influences the measure outcome.

Multivariable model refers to statistical models that examine relationships among more than two variables. A multivariable model can be thought of as a model in which multiple variables are found on the right side of the model equation. This type of statistical model can be used to attempt to assess the

relationship between a number of variables; one can assess independent relationships while adjusting for potential confounders. A multivariable model, therefore, contains more than one predictor to predict that single outcome.

Proxy factors refer to any correlate of a strong risk factor that may also appear to be a risk factor for the same outcome, even though the only connection between that correlate and the outcome lies in the strong risk factor correlated with both.[79]

Outcome is used broadly to refer to the results of care delivery, which include the following types of outcomes relevant to performance measurement: health outcomes (e.g., mortality, adverse events), intermediate clinical outcome (e.g., BP < 140/90), economic outcomes of cost and resource use, and patient-reported outcomes (e.g., symptoms, mood).

Overfitting describes risk adjustment models that contain too many variables such that they begin to describe noise or qualities of the data set rather than an underlying relationship between the intervention and outcome. There are a variety of statistical techniques to reduce the number of variables in the model due to overfitting.

Quality of care refers to a measure of performance on the six IOM-specified healthcare aims: (1) safety, (2) timeliness, (3) effectiveness, (4) efficiency, (5) equity, and (6) patient-centeredness.[18]

Reliability refers to the ability to yield consistent and reproducible results. Statisticians call this characteristic precision, whereas social scientists, psychologists, and health services researchers know it as *reliability*.[14]

Risk adjustment (also known as case-mix adjustment) refers to statistical methods to control or account for patient- and/or community-level factors when computing performance measure scores; methods include modeling techniques, indirect standardization, or direct standardization. These methods can be used to produce a ratio of observed-to-expected, a risk-adjusted rate, or another estimate of performance. Methods include, but are not limited to, adjustment for mean within-reporting unit differences in multivariable models with reporting unit fixed effects, indirect standardization, direct standardization, and matched cohort comparisons.[14]

Social drivers of health (SDOH) (also known as social determinants of health) are the social, nonmedical conditions that determine healthcare provision and health outcomes.[28] They can both improve and worsen an individual's health. SDOH and social risk factors are connected: SDOH can impact a person's health for better or worse, depending on social circumstances. However, when those social circumstances are adverse, then some people may be at greater risk for poor health. These circumstances are referred to as social risk factors.

Social risk adjustment refers to statistical adjustment for social variables, including those that are socioeconomic (e.g., income, education, occupation) and sociodemographic (e.g., age, race, gender, gender identity, Medicare/Medicaid dual eligibility, language).

Social risk factors are broadly defined to include the social conditions or factors that have a conceptual and empirical relationship to healthcare outcomes.[28] Illustratively, these factors may include socioeconomic position or status (e.g., income, education, occupation), other cultural context, social relationships, residential and community environments, urbanicity/rurality, and health literacy. Consistent with the National Academies of Sciences, Engineering and Medicine (NASEM) and Assistant Secretary for Planning and Evaluation (ASPE) reports,[27] this guidance considers a variety of sociodemographic factors as social risk factors, including age, gender, gender identity, Medicare/Medicaid dual eligibility, language, and uninsured status. Throughout this guidance, age is treated as both a clinical and social risk factor.[28]

Stratification refers to an approach to identifying disparities. In addition to reporting overall performance, stratification consists of computing performance separately for different strata or groupings of patients based on some patient-level characteristic(s), namely those that are social and functional status-related for the purposes of this guidance. Thus, each accountable entity has multiple performance scores, one for each stratum rather than one overall performance score.[7]

Validity shows how well the adjustment method accounts for the true risk of a specified outcome within a particular time frame for a particular patient population for a specific purpose.[14]

Value-based purchasing (VBP) refers to a wide variety of payment strategies that incentivize providers to deliver high value healthcare by linking provider performance and quality of care with payment incentives.

Within-unit differences occur when quality varies across different providers or units within a measured entity, regardless of accountable entities' case mix. For example, a hospital that provides lower quality care only for socially disadvantaged patients is exhibiting within-unit differences.[68,80]

Appendix C: Social and Functional Risk Data Sources

Data Source	Strengths	Limitations
Administrative Claims	• Useful for tracking healthcare resource utilization and cost-related information • Range of data includes anything that is reimbursed by health insurance, generally including visits to physicians and allied health providers, most prescription drugs, many devices, hospitalization(s) (if a lab test was performed), and in some cases, actual lab test results for selected tests (e.g., blood test results for cholesterol, diabetes). • In some cases, demographic information (e.g., gender, date of birth from billing files) can be available. • Potential for efficient capture of large populations	• Represent clinical cost drivers versus complete clinical diagnostic and treatment information • It is important to be knowledgeable about the process and standards used in claims submission. For example, only a primary diagnosis may be coded and secondary diagnoses not captured. In other situations, value-laden claims may not be used (e.g., an event may be coded as a "nonspecific gynecologic infection" rather than a "sexually transmitted disease"). • Important to be knowledgeable about data handling and coding systems used when incorporating the claims data into the administrative systems • Can be difficult to gain the cooperation of partner groups, particularly with regard to receiving the submissions in a timely manner • May be limited to specific demographics, such as 65+ Medicare beneficiaries

Data Source	Strengths	Limitations
Electronic Health Records (EHRs)	• Information on routine medical care and practice, with more clinical context than coded claims • Potential for comprehensive view of patient medical and clinical history • Efficient access to medical and clinical data • Use of data transfer and coding standards (including handling of missing data) will increase the quality of data abstracted	• Underlying information from clinicians is not collected using uniform decision rules. (See example under "Medical chart abstraction.") • Consistency of data quality and breadth of data collected varies across sites • Difficult to handle information uploaded as text files into the EHRs (e.g., scanned clinician reports) versus direct entry into data fields • Historical data capture may require manual chart abstraction prior to implementation date of medical records system. • Complete medical and clinical history may not be available (e.g., new patient to clinic). • EHR systems vary widely. If data come from multiple systems, the registry should plan to work with each system individually to understand the requirements of the transfer.
Registry Data	• Generally, the most granular, standardized clinical data available • Typically entered by trained coders • All payers and ages • Can be merged with another data source to answer additional questions not considered in the original registry protocol or plan • May include specific data not generally collected in routine medical practice • Can provide historical comparison data	• Increased data collection burden and cost • May be limited to one disease process or procedure • Important to understand the existing registry protocol or plan to evaluate data collected for element definitions, timing, and format, as it may not be possible to merge data unless many of these aspects are similar.

Data Source	Strengths	Limitations
Clinical Assessment Data	• Patient and/or caregiver outcomes • Unique perspective • Obtain information on treatments not necessarily prescribed by clinicians (e.g., over-the-counter drugs, herbal medications) • Obtain intended compliance information • Useful when timing of follow-up may not be concordant with timing of clinical encounter	• Literacy, language, or other barriers that may lead to under-enrollment of some subgroups • Validated data collection instruments may need to be developed. • Loss to follow-up or refusal to continue participation • Limited confidence in reporting clinical information and utilization information

Appendix D: Examples of Approaches to Social and/or Functional Risk Adjustment

The Technical Guidance in part reflects current and recent approaches to assessing social risk variables presented in measure endorsement submissions to National Quality Forum (NQF). To provide examples of how developers could meet certain components of the Technical Guidance, this appendix shares example content from submitted measure forms that aligns with the Technical Guidance. The examples presented below are verbatim figures, tables, and text extracted from performance measure endorsement submissions 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**.

Below we note the relevant portion of the Technical Guidance that each example illustrates, then provide the verbatim examples.

Examples That Illustrate Conceptualizing the Model

Below, Examples 1, 2, and 3 are verbatim text submitted by measure developers that ground the selection of risk factors in a conceptual model as envisioned by the Technical Guidance.

- Example 1 depicts a developer's explanation of a conceptual model, including the literature review and caveats up front for the hypotheses made during the conceptualization of the model.
- Example 2 shows a conceptual model graphic accompanied by a narrative description of its contents.
- Example 3 depicts how a developer examined the literature to identify social risk factors for the conceptual model and subsequently used the conceptual model to guide the selection of variables (both clinical and social) for empirical examination.

Example 1. NQF #2880 Excess Days in Acute Care (EDAC) After Hospitalization for Heart Failure (HF) – NQF-Endorsed (Yale CORE / Centers for Medicare & Medicaid Services)

Conceptual Model for Risk Adjustment:

Our approach to risk adjustment is tailored to, and appropriate for, a publicly reported outcome measure as articulated in published scientific guidelines (Krumholz et al, 2006, Normand et al, 2007). We adopted the risk factors from the existing NQF-endorsed, Centers for Medicare & Medicaid Services' (CMS) measure titled *30-Day Heart Failure (HF) Readmission* (Dorsey et al 2015). These risk factors comprise age, sex, and condition categories (CCs) for prior 12-month and current claims. These risk factors had been systematically chosen as predictors of any readmission for the same patient cohort as the current measure; the outcome of this measure is dominated by the number of days of a readmission, so we judged it unlikely that repeating the original analysis would produce different results. We confirmed that there were no additional risk factors to consider by comparing the model estimated using the *a priori* set of risk factors to a model, which included all additional CCs.

For risk adjustment, we used a hierarchical generalized linear model (HGLM). The model consists of two parts: a logit model and a truncated Poisson model. The two-part logit/Poisson model (often called a *hurdle model)* assumes that the outcome results from two related processes: (1) an initial dichotomous event, assuming that a patient has at least one acute care event, which is modeled as the logit of the probability of the event and (2) for patients with an event (those who clear the "hurdle"), the number of days, which is modeled as a Poisson process. The outcome, which is the number of days, is a half-integer count variable (because ED visits count as 0.5 days). Observation care is counted according to the hours spent in observation care rounded up to the nearest half-day. For each patient, an exposure variable is defined as the number of survival days post-discharge up to 30. For the hurdle model, exposure time as an offset is included for each part of the model.

There are two random effects for each hospital: one for the logit model and one for the truncated Poisson model, as well as a covariance between the two random effects. The random effects allow us to account for within-hospital correlation of the observed outcome and accommodate the assumption that underlying differences in quality across hospitals lead to systematic differences in outcomes.

Socioeconomic Status Factors and Race

We selected variables representing SES factors and race for examination based on a review of literature, conceptual pathways, and feasibility. In Section 1.8, we describe the variables that we considered and analyzed based on this review. Below, we describe the pathways by which SES and race may influence days in acute care in the 30 days after discharge.

Our conceptualization of the pathways by which patient SES or race affects days in acute care within the 30 days is informed by the literature on the association of SES and race with heart failure (HF) readmissions since the majority of the Excess Days in Acute Care (EDAC) outcome is composed of readmission days and considering that there is much more robust literature about readmission than observation care and emergency department (ED) visits.

Literature Review of Socioeconomic Status and Race Variables and Heart Failure Excess Days in Acute Care

To examine the relationship between SES and race variables and hospital 30-day, all-cause EDAC following HF hospitalization, a literature search was performed with the following exclusion criteria: international studies, articles published more than 10 years ago, articles without primary data, articles using Veterans Affairs (VA) databases as the primary data source, and articles not explicitly focused on SES or race and HF readmission. Fifty studies were initially reviewed, and 36 studies were excluded from a full-text review based on the above criteria. Studies indicated that SES/race variables were associated with increased risk of (HF) readmission (Foraker et al, 2011; Kind et al, 2014; Vivo et al, 2014; Joynt, Orav, and Jha 2011; Lindenauer et al, 2013; Allen et al, 2012; Regalbuto et al, 2014; Aseltine et al, 2015; Calvillo-King et al, 2013; McHugh, Carthon, and Kang 2010; Damiani et al, 2015; Berenson and Shih 2012), although there may not be a significant effect on hospital-level profiling (Blum et al, 2014).

Causal Pathways for Socioeconomic Status and Race Variable Selection

Although some recent literature evaluates the relationship between patient SES or race and the readmission outcome, few studies directly address causal pathways or examine the role of the hospital in these pathways. Moreover, the current literature examines a wide range of conditions and risk variables with no clear consensus on which risk factors demonstrate the strongest relationship with readmission. The SES factors that have been examined in the readmission literature can be categorized into three domains: (1) patient-level variables, (2) neighborhood/community-level variables, and (3) hospital-level variables. Patient-level variables describe characteristics of individual patients and range from the self-reported or documented race or ethnicity of the patient to the patient's income or education level (Eapen et al, 2015; Hu et al, 2014). Neighborhood/community-level variables use information from sources such as the ACS as either a proxy for individual patient-level data or to measure environmental factors. Studies using these variables use one-dimensional measures, such as median household income or composite measures, such as the Agency for Healthcare Research and Quality (AHRQ)-validated SES index score (Blum et al, 2014). Hospital-level variables measure attributes of the hospital, which may be related to patient risk. Examples of hospital-level variables used in studies are ZIP code characteristics aggregated to the hospital level or the proportion of Medicaid patients served in the hospital (Gilman et al, 2014; Joynt and Jha 2013).

The conceptual relationship, or potential causal pathways by which these possible SES risk factors influence the risk of readmission following an acute illness or major surgery, such as the factors themselves, are both varied and complex. There are at least four potential pathways that are important to consider:

1. **Relationship of SES factors or race to health at admission**. Patients who have lower income/education/literacy or unstable housing may have a worse general health status and may present for their hospitalization or procedure with a greater severity of underlying illness. These SES risk factors, which are characterized by patient-level or neighborhood/community-level (as proxy for patient-level) variables, may contribute to a worse health status at admission due to competing priorities (e.g., restrictions based on job, lack of childcare), lack of access to care (e.g., geographic, cultural, or financial), or lack of health insurance. Given that these risk factors all lead to worse general health status, this causal pathway should be largely accounted for by current clinical risk adjustment.

   In addition to SES risk factors, studies have shown that worse health status is more prevalent among African American patients compared with White patients. The association between race and worse health is in part mediated by the association between race and SES risk factors, such as poverty or disparate access to care associated with poverty or neighborhood. The association is also mediated through bias in healthcare as well as in other facets of society.

2. **Use of low-quality hospitals**. Patients of lower income, lower education, or unstable housing have been shown not to have equitable access to high quality facilities because such facilities are less likely to be found in geographic areas with large populations of poor patients; thus, patients with low income are more likely to be seen in lower-quality hospitals, which can contribute to increased risk of readmission following hospitalization (Jha et al, 2011; Reames et al, 2014). Similarly, African

American patients have been shown to have less access to high quality facilities compared with White patients (Skinner et al, 2005).

3. **Differential care within a hospital**. The third major pathway by which SES factors or race may contribute to readmission risk is patients who may not receive equivalent care within a facility. For example, African American patients have been shown to experience differential, lower quality, or discriminatory care within a given facility (Trivedi et al, 2014). Alternatively, patients with SES risk factors, such as lower education, may require differentiated care (e.g., provision of lower literacy information) that they do not receive.

4. **Influence of SES on readmission risk outside of hospital quality and health status**. Some SES risk factors, such as income or wealth, may affect the likelihood of readmission without directly affecting health status at admission or the quality of care received during the hospital stay. For instance, while a hospital may make appropriate care decisions and provide tailored care and education, a lower-income patient may have a worse outcome post-discharge due to competing economic priorities or a lack of access to care outside of the hospital.

These proposed pathways are complex to distinguish analytically. They also have different implications on the decision to risk-adjust or not. Therefore, we first assessed whether there was sufficient evidence of a meaningful effect on the risk model to warrant efforts to distinguish among these pathways. Based on this model and the considerations outlined in Section 1.8, the following SES and race variables were considered:

- Dual-eligible status
- African American race

We assessed the relationship between the dual-eligible status and race with the outcome and examined the incremental effect of each in a multivariable model. For this measure, we also examined the extent to which the addition of any one of these variables improved model performance or changed hospital results.

One concern with including SES or race factors in a model is that their effect may be at either the patient or the hospital level. For example, low SES may increase the risk of readmission because patients of low SES have a higher individual risk (patient-level effect), or because patients of low SES are more often admitted to hospitals with higher overall readmission rates (hospital-level effect). Thus, as an additional step, we performed a decomposition analysis to assess the independent effects of the SES and race variables at the patient level and hospital levels. If, for example, all the elevated risk of readmission for patients of low SES was due to lower-quality/higher readmission risk in hospitals with more patients of low SES, then a significant hospital-level effect would be expected with little-to-no patient-level effect. However, if the increased readmission risk was solely related to higher risk for patients of low SES regardless of hospital effect, then a significant patient-level effect would be expected, and a significant hospital-level effect would not be expected.

Specifically, we decomposed each of the SES and race variables as follows: Let *Xij* be a binary indicator of the SES or race status of the *ith* patient at the *jth* hospital and *Xj* be the percent of patients at hospital *j* with Xij = 1. Then, we rewrote Xij = (Xij- Xj) + Xj ≡ Xpatient+ Xhospital. The first variable, *Xpatient*, represents the effect of the risk factor at the patient level (sometimes called the *within hospital effect*), and the second variable, *Xhospital*, represents the effect at the hospital level (sometimes called the *between hospital effect*). By including both variables in the same model, we can assess whether these are independent effects or whether only one of these effects contributes. This analysis allows us to simultaneously estimate the independent effects of these two classifying groups: (1) hospitals with higher or lower proportions of low SES patients or African American patients on the readmission rate of an average patient and (2) a patient's SES or race on their own readmission rates when seen at an average hospital.

It is very important to note, however, that even in the presence of a significant patient-level effect and absence of a significant hospital-level effect, the increased risk could be partly or entirely due to the quality of care patients receive in the hospital. For example, biased or differential care provided within a hospital to low-income patients compared with high-income patients would exert its impact at the level of individual patients and would therefore be a patient-level effect. It is also important to note that the patient-level and hospital-level coefficients cannot be quantitatively compared because the patient's SES circumstance or race in the model is binary, whereas the hospitals' proportion of low SES patients or African American patients is continuous.

References

Allen LA, Smoyer Tomic KE, Smith DM, Wilson KL, Agodoa I. Rates and predictors of 30-day readmission among commercially insured and Medicaid-enrolled patients hospitalized with systolic heart failure. Circulation. Heart failure. 2012;5(6):672-679.

Aseltine RH, Jr., Yan J, Gruss CB, Wagner C, Katz M. Connecticut Hospital Readmissions Related to Chest Pain and Heart Failure: Differences by Race, Ethnicity, and Payer. Connecticut medicine. 2015;79(2):69-76.

Berenson J, Shih A. Higher readmissions at safety-net hospitals and potential policy solutions. Issue brief (Commonwealth Fund). 2012;34:1-16.

Blum, A. B., N. N. Egorova, E. A. Sosunov, A. C. Gelijns, E. DuPree, A. J. Moskowitz, A. D. Federman, D. D. Ascheim and S. Keyhani. "Impact of Socioeconomic Status Measures on Hospital Profiling in New York City." Circ Cardiovasc Qual Outcomes 7, no. 3 (2014): 391-7.

Calvillo-King L, Arnold D, Eubank KJ, et al. Impact of social factors on risk of readmission or mortality in pneumonia and heart failure: systematic review. Journal of general internal medicine. 2013;28(2):269-282.

Damiani G, Salvatori E, Silvestrini G, et al. Influence of socioeconomic factors on hospital readmissions for heart failure and acute myocardial infarction in patients 65 years and older: evidence from a systematic review. Clinical interventions in aging. 2015;10:237-245.

Dorsey KB GJ, Desai N, Lindenauer P, Young J, Wang C, DeBuhr, Parisi ML, Bernheim SM, Krumholz HM. 2015 Condition-Specific Measures Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Readmission Measures: AMI-Version 8.0, HF-Version 8.0, Pneumonia-Version 8.0, COPD-Version 4.0, and Stroke-Version 4.0. 2015; https://www.qualitynet.org/dcs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1228890435217&blobheader=multipart%2Foctet-stream&blobheadername1=Content-Disposition&blobheadervalue1=attachment%3Bfilename%3DRdmn_AMIHFPNCOPDSTK_Msr_UpdtRpt.pdf&blobcol=urldata&blobtable=MungoBlobs. Accessed July 8, 2015.

Eapen ZJ, McCoy LA, Fonarow GC, Yancy CW, Miranda ML, Peterson ED, Califf RM, Hernandez AF. Utility of socioeconomic status in predicting 30-day outcomes after heart failure hospitalization. Circ Heart Fail. May 2015; 8(3):473-80.

Foraker, R. E., K. M. Rose, C. M. Suchindran, P. P. Chang, A. M. McNeill and W. D. Rosamond. "Socioeconomic Status, Medicaid Coverage, Clinical Comorbidity, and Rehospitalization or Death after an Incident Heart Failure Hospitalization: Atherosclerosis Risk in Communities Cohort (1987 to 2004)." Circ Heart Fail 4, no. 3 (2011): 308-16.

Gilman M, Adams EK, Hockenberry JM, Wilson IB, Milstein AS, Becker ER. California safety-net hospitals likely to be penalized by ACA value, readmission, and meaningful-use programs. Health Aff (Millwood). Aug 2014; 33(8):1314-22.

Hu J, Gonsahn MD, Nerenz DR. Socioeconomic status and readmissions: evidence from an urban teaching hospital. Health affairs (Project Hope). 2014; 33(5):778-785.

Jha AK, Orav EJ, Epstein AM. Low-quality, high-cost hospitals, mainly in South, care for sharply higher shares of elderly black, Hispanic, and Medicaid patients. Health Affairs 2011; 30:1904-11.

Joynt KE, Jha AK. Characteristics of hospitals receiving penalties under the Hospital Readmissions Reduction Program. JAMA. Jan 23 2013; 309(4):342-3.

Joynt, K. E., E. J. Orav and A. K. Jha. "Thirty-Day Readmission Rates for Medicare Beneficiaries by Race and Site of Care." JAMA 305, no. 7 (2011): 675-81.

Kind, A. J., S. Jencks, J. Brock, M. Yu, C. Bartels, W. Ehlenbach, C. Greenberg and M. Smith. "Neighborhood Socioeconomic Disadvantage and 30-Day Rehospitalization: A Retrospective Cohort Study." Ann Intern Med 161, no. 11 (2014): 765-74.

Krumholz HM, Brindis RG, Brush JE, et al. 2006. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on

Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. Circulation 113: 456-462.

Lindenauer, P. K., T. Lagu, M. B. Rothberg, J. Avrunin, P. S. Pekow, Y. Wang and H. M. Krumholz. "Income Inequality and 30 Day Outcomes after Acute Myocardial Infarction, Heart Failure, and Pneumonia: Retrospective Cohort Study." Bmj 346, (2013): f521.

McHugh MD, Carthon JM, Kang XL. Medicare readmissions policies and racial and ethnic health disparities: a cautionary tale. Policy, politics & nursing practice. 2010;11(4):309-316.

Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. Stat Sci 22 (2): 206-226.

Reames BN, Birkmeyer NJ, Dimick JB, Ghaferi AA. Socioeconomic disparities in mortality after cancer surgery: failure to rescue. JAMA surgery 2014; 149:475-81.

Regalbuto R, Maurer MS, Chapel D, Mendez J, Shaffer JA. Joint Commission requirements for discharge instructions in patients with heart failure: is understanding important for preventing readmissions? Journal of cardiac failure. 2014;20(9):641-649.

Skinner J, Chandra A, Staiger D, Lee J, McClellan M. Mortality after acute myocardial infarction in hospitals that disproportionately treat black patients. Circulation 2005; 112:2634-41.

Trivedi AN, Nsa W, Hausmann LR, et al. Quality and equity of care in U.S. hospitals. The New England journal of medicine 2014; 371:2298-308.

Vivo, R. P., S. R. Krim, L. Liang, M. Neely, A. F. Hernandez, Z. J. Eapen, E. D. Peterson, D. L. Bhatt, P. A. Heidenreich, C. W. Yancy and G. C. Fonarow. "Short- and Long-Term Rehospitalization and Mortality for Heart Failure in 4 Racial/Ethnic Populations." J Am Heart Assoc 3, no. 5 (2014): e001134.

Example 2. NQF #3597 Clinician-Group Risk-Standardized Acute Hospital Admission Rate for Patients With Multiple Chronic Conditions Under the MIPS – NQF-Endorsed (Yale CORE / Centers for Medicare & Medicaid Services)

Conceptual Model for Risk Adjustment:

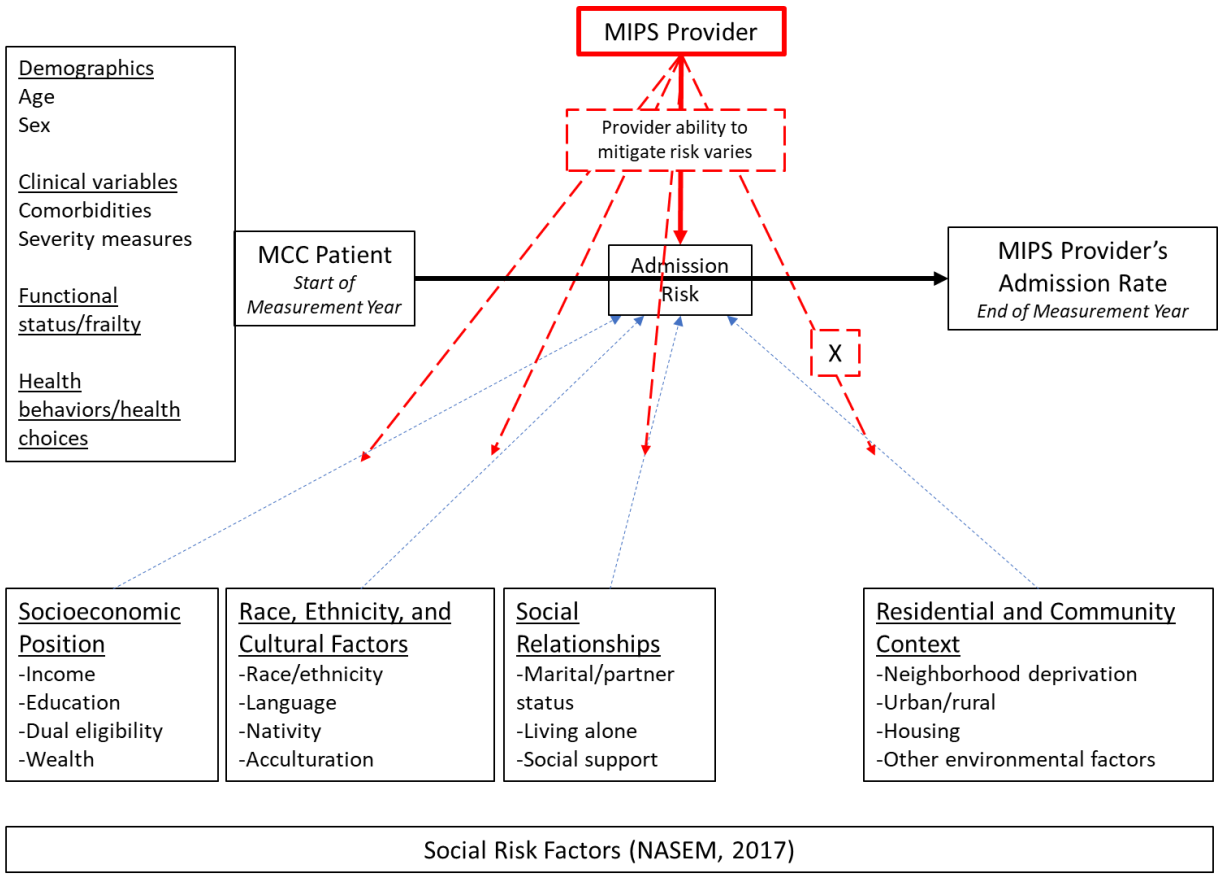
The MIPS Multiple Chronic Conditions (MCC) measure is built as an adaptation of a similar measure developed for CMS that identifies acute admission rates for MCC patients in the Accountable Care Organization (ACO) setting [2]. Building on the conceptual model developed in that measure, we defined and illustrated the potential relationships between different categories of risk factors and the outcome of hospital admissions. This MIPS conceptual model (see the figure below) guided the selection of candidate risk factors. We identified patient demographic factors and clinical variables, including comorbidities and measures of frailty and disability, which reflect the characteristics of the patients at the start of the measurement year and are independent of the quality of care. The potential clinical variables included not only clinical comorbidities but also measures of disease severity and frailty/functional status.

We also considered social risk factors that may influence patients' risk of acute, unplanned admissions. There are many ways to conceptualize or categorize social risk factors. We adopted the model of the National Academies of Sciences, Engineering, and Medicine's (NASEM) comprehensive, expert report of 2017, in which they categorized social risk factors into the following four domains [3]:

1. Socioeconomic position
2. Race, ethnicity, and cultural factors
3. Social relationships
4. Residential and community context

(Note: There is a fifth domain in the NASEM report related to gender and sexual orientation; however, we have omitted it because the authors noted that more research is needed to understand the relationship between these factors and outcomes and because of a lack of available data.)

Figure 1: Conceptual Model for Risk Adjustment



As noted in our conceptual model (Figure 1), variables in all of these domains are to be or are hypothesized to be associated with increased risk of admission. However, the domains differ in the extent to which we expect an individual MIPS clinician or group of clinicians to be able to mitigate the risk conferred by such variables. These differences inform their potential use as risk-adjusters since adjusting for factors that can be more easily mitigated by higher quality care is more likely to mask low-quality care.

MIPS providers have the least ability to mitigate the risk of admission associated with broader residential and community factors, such as neighborhood deprivation and relative lack of access to primary and specialty medical care. In contrast, we expect that there is more, although limited, ability for a MIPS provider to intervene to mitigate some or all of the risk conferred by the other individual-level domains noted above. For example, a provider can consider a patient's education level, health literacy level, and home living situation when planning and delivering care. In addition, high quality care may be characterized as being more racially, linguistically, and culturally sensitive and informed. While such tailored care can likely mitigate the risk of admission, our TEP emphasized that providing it also requires resources; as a result, MIPS providers may be limited in their capacity to deliver it.

[1] Krumholz HM, Brindis RG, Brush JE, et al. 2006. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. Circulation 113: 456-462.

[2] Centers for Medicare & Medicaid Services. #2888 Risk-Standardized Acute Admission Rates for Patients with Multiple Chronic Conditions, Last Updated: Jan 09, 2019. 2016; http://www.qualityforum.org/ProjectTemplateDownload.aspx?SubmissionID=2888. Accessed February 20, 2019.

[3] Steinwachs DM, Stratton, K., Kwan, L. Y.,. Accounting for Social Risk Factors in Medicare Payment. Washington DC: 2017 by the National Academy of Sciences; 2017.

Example 3. NQF #1789 Hospital-Wide Readmission Measure (HWR) – NQF-Endorsed (Yale CORE / Centers for Medicare & Medicaid Services)

Approach to Variable Selection:

In order to select the comorbid risk variables, we developed a "starter" set of 30 variables drawn from previous readmission measures (e.g., acute myocardial infarction [AMI], HF, pneumonia, hip and knee arthroplasty, and stroke). Next, we reviewed all the remaining CMS-CCs and determined on a clinical basis whether they were likely to be relevant to an all-condition measure. We selected 11 additional risk variables for consideration.

Using data from the index admission and any admission in the prior 12 months, we ran a standard logistic regression model for every discharge condition category with the full set of candidate risk adjustment variables. We compared odds ratios for different variables across different condition categories (excluding condition categories with fewer than 700 readmissions due to the number of events per variable constraints). We selected the final set of comorbid risk variables based on the following principles:

- We excluded risk variables that were statistically significant for very few condition categories, given that they would not contribute much to the overall models.
- We excluded risk variables that behaved in clinically incoherent ways. For example, we dropped risk variables that at times increased risk and at other times decreased risk when we could not identify a clinical rationale for the differences.
- We excluded risk variables that were predominantly protective when we felt this protective effect was not clinically reasonable but more likely reflected coding factors. For example, drug/alcohol abuse without dependence (CC 53) and delirium and encephalopathy (CC 48) were both protective for readmission risk, although clinically they should increase patients' severity of illness.
- Where possible, we grouped together risk variables that were clinically coherent and carried similar risks across condition categories. For example, we combined coronary artery disease (CCs 83-84) with cerebrovascular disease (CCs 98, 99, and 103).
- We examined risk variables that had been combined in previous CMS publicly reported measures, and in one instance, we separated them: For cancers, the previous measures generally pool five categories of cancers (CCs 8 to 12) together. In our analysis, lung cancer (CC 8) and other severe cancers (CC 9) carried higher risks, so we separated them into a distinct risk variable and grouped other major cancers (CC 10), benign cancers (CC 11), and cancers of the urinary and gastrointestinal (GI) tracts (CC 12) together. Consistent with other publicly reported measures, we also left metastatic cancer/leukemia (CC 7) as a separate risk variable.

Complications occurring during hospitalization are not comorbid illnesses and may reflect the hospital's quality of care; therefore, they should not be used for risk adjustment. Hence, conditions that may represent adverse outcomes due to the care received during the index hospital stay are not included in the risk-adjusted model (see Table 5 in Section 2a1.13). CCs on this list were not counted as a risk variable in our analyses if they appeared only on the index admission.

Service Mix Adjustment:

The measure includes many different discharge condition categories that differ in their baseline readmission risks. In addition, hospitals differ in their relative distribution of these condition categories (i.e., service mix). To adjust for service mix, the measure uses an indicator variable for the discharge condition category, in addition to risk variables for comorbid conditions. The models include the following items:

- A condition-specific indicator for all-condition categories with sufficient volume (defined as those with more than 1,000 admissions nationally in a given year for Medicare FFS data) as well as a single indicator for conditions with insufficient volume in each model
- SES factors and race
- SES factors and race for examination were based on a review of literature, conceptual pathways, and feasibility. In Section 1.8, we describe the variables that we considered and analyzed based on this review. Below, we describe the pathways by which SES and race may influence 30-day readmission.
- Our conceptualization of the pathways by which patient SES or race affects 30-day readmission is informed by the literature.
- SES and race variables and Hospital Wide Readmission (HWR)

To examine the relationship between SES, race variables, and hospital 30-day, hospital-wide, all-cause, unplanned readmission following hospitalization, a literature search was performed with the following exclusion criteria: international studies, articles published more than 10 years ago, articles without primary data, articles using VA databases as the primary data source, and articles not explicitly focused on SES or race and readmission across multiple conditions. One hundred sixty-nine articles were initially reviewed, and 155 studies were excluded from a full-text review based on the above criteria. Studies indicate that SES/race variables were associated with increased risk of readmission across multiple major illnesses and conditions (Aseltine RH, et al, 2015; Mitchell SE, et al, 2012; Odonkor CA, et al, 2015; Herrin J, et al, 2015; Gu Q, et al, 2014, Kim H, et al, 2010; Kangovi S, et al, 2012; Iloabuchi TC, 2014; Beck AF, et al, 2012; Arbaje AI, et al, 2008; Hu J, 2014; Nagasako EM, et al, 2014; Joynt, KE, et al, 2013), although there may not be a significant effect on hospital-level profiling (Blum AB, et al, 2014).

SES and Race Variable Selection:

Although some recent literature evaluates the relationship between patient SES or race and the readmission outcome, few studies directly address causal pathways or examine the role of the hospital in these pathways. Moreover, the current literature examines a wide range of conditions and risk variables with no clear consensus on which risk factors demonstrate the strongest relationship with readmission. The SES factors that have been examined in the readmission literature can be categorized into three domains: (1) patient-level variables, (2) neighborhood/community-level variables, and (3) hospital-level variables. Patient-level variables describe characteristics of individual patients and range from the self-reported or documented race or ethnicity of the patient to the patient's income or education level (Eapen ZJ, et al, 2015; Hu J, et al, 2014). Neighborhood/community-level variables use information from sources such as the ACS as either a proxy for individual patient-level data or a tool to

measure environmental factors. Studies using these variables use one-dimensional measures, such as median household income or composite measures, such as the AHRQ-validated SES index score (Blum AB, et al, 2014). Hospital-level variables measure attributes of the hospital, which may be related to patient risk. Examples of hospital-level variables used in studies are ZIP code characteristics aggregated to the hospital level or the proportion of Medicaid patients served in the hospital (Gilman M, et al, 2014; Joynt KE and Jha AK, 2013).

The conceptual relationship and the potential causal pathways by which these possible SES risk factors and race/ethnicity influence the risk of readmission following an acute illness or major surgery, such as the factors themselves, are both varied and complex. There are at least four potential pathways that are important to consider:

1. **Relationship of SES factors or race to health at admission**. Patients who have lower income/education/literacy or unstable housing may have a worse general health status and may present for their hospitalization or procedure with a greater severity of underlying illness. These SES risk factors, which are characterized by patient-level or neighborhood/community-level (as proxy for patient-level) variables, may contribute to a worse health status at admission due to competing priorities (e.g., restrictions based on job, lack of childcare), lack of access to care (e.g., geographic, cultural, or financial), or lack of health insurance. Given that these risk factors all lead to worse general health status, this causal pathway should be largely accounted for by current clinical risk adjustment.

   In addition to SES risk factors, studies have shown that worse health status is more prevalent among African American patients compared with White patients. The association between race and worse health is in part mediated by the association between race and SES risk factors, such as poverty or disparate access to care associated with poverty or neighborhood. The association is also mediated through bias in healthcare as well as other facets of society.

2. **Use of low-quality hospitals**. Patients of lower income, lower education, or unstable housing have been shown not to have equitable access to high quality facilities because such facilities are less likely to be found in geographic areas with large populations of poor patients; thus, patients with low income are more likely to be seen in lower quality hospitals, which can contribute to increased risk of readmission following hospitalization (Jha AK, et al, 2011; Reames BN, et al, 2014). Similarly, African American patients have been shown to have less access to high quality facilities compared with White patients (Skinner J, et al., 2005).

3. **Differential care within a hospital**. The third major pathway by which SES factors or race may contribute to readmission risk is patients who may not receive equivalent care within a facility. For example, African American patients have been shown to experience differential, lower-quality, or discriminatory care within a given facility (Trivedi AN, et al, 2014). Alternatively, patients with SES risk factors, such as lower education, may require differentiated care (e.g., provision of lower literacy information) that they do not receive.

4. **Influence of SES on readmission risk outside of hospital quality and health status**. Some SES risk factors, such as income or wealth, may affect the likelihood of readmission without directly affecting health status at admission or the quality of care received during the hospital stay. For instance, while a hospital may make appropriate care decisions and provide tailored care and education, a lower-income patient may have a worse outcome post-discharge due to competing economic priorities or a lack of access to care outside of the hospital.

These proposed pathways are complex to distinguish analytically. They also have different implications on the decision to risk-adjust or not. Therefore, we first assessed whether there was evidence of a meaningful effect on the risk model to warrant efforts to distinguish among these pathways. Based on this model and the considerations outlined in Section 1.8, the following SES and race variables were considered:

- Dual-eligible status
- African American race
- AHRQ SES index

We assessed the relationship between the SES variables and race with the outcome and examined the incremental effect in a multivariable model. For this measure, we also examined the extent to which the addition of any one of these variables improved model performance or changed hospital results.

One concern with including SES or race factors in a model is that their effect may be at either the patient or the hospital level. For example, low SES may increase the risk of readmission because patients of low SES have a higher individual risk (patient-level effect) or because patients of low SES are more often admitted to hospitals with higher overall readmission rates (hospital-level effect). Thus, as an additional step, we performed a decomposition analysis to assess the independent effects of the SES and race variables at the patient and hospital levels. If, for example, all the elevated risk of readmission for patients of low SES was due to lower-quality/higher-readmission risk in hospitals with more patients of low SES, then a significant hospital-level effect would be expected with little-to-no patient-level effect. However, if the increased readmission risk was solely related to higher risk for patients of low SES regardless of hospital effect, then a significant patient-level effect would be expected, and a significant hospital-level effect would not be expected.

Specifically, we decomposed each of the SES and race variables as follows: Let Xij be a binary indicator of the SES or race status of the ith patient at the jth hospital and Xj be the percent of patients at hospital j with Xij = 1. Then, we rewrote Xij = (Xij- Xj) + Xj $\equiv$ Xpatient+ Xhospital. The first variable, Xpatient, represents the effect of the risk factor at the patient level (sometimes called the *within hospital effect*), and the second, Xhospital, variable represents the effect at the hospital level (sometimes called the *between hospital effect*). By including both variables in the same model, we can assess whether these are independent effects or whether only one of these effects contributes. This analysis allows us to simultaneously estimate the independent effects of these two classifying groups: (1) hospitals with higher or lower proportions of low SES patients or African American patients on the readmission rate of an average patient and (2) a patient's SES or race on their own readmission rates when seen at an average hospital.

It is very important to note, however, that even in the presence of a significant patient-level effect and absence of a significant hospital-level effect, the increased risk could be partly or entirely due to the quality of care patients receive in the hospital. For example, biased or differential care provided within a hospital to low-income patients as compared with high-income patients would exert its impact at the level of individual patients and would therefore be a patient-level effect. It is also important to note that the patient-level and hospital-level coefficients cannot be quantitatively compared because the patient's SES circumstance or race in the model is binary, whereas the hospitals' proportion of low SES patients or African American patients is continuous.

Accountable Care Organization:

In considering the modification of this measure for the ACO program, we were guided by a conceptual framework outlining the relationships between potential, clinical, and contextual factors and rates of readmission at the ACO level. Importantly, many factors other than traditional medical care delivered in the office or hospital settings will have an impact on the likelihood of readmission. For example, ACOs practicing in communities where patients have limited access to transportation, healthy foods, and recreational facilities may have less success in promoting healthy behaviors among patients; this may, in turn, have an impact on readmission rates. Recognition of and attention to the health environment may be important for achieving the goals of better care, better health, lower costs, and thus, shared savings.

Our conceptual model recognizes patient-level demographic and clinical factors, along with four contextual domains that may influence ACO performance: (1) physical environment (e.g., green spaces, safe streets); (2) community resources (e.g., home health, senior services); (3) patient resources (e.g., social support, transportation, and income); and (4) patient behavior/personal preferences (e.g., exercise, diet, advanced care directives, and preference for intervention).

The model also recognizes the capacity of ACOs to mitigate the effects of many contextual factors on rates of admissions, encompassing both SES and non-SES variables. Adjusting for contextual factors would obscure important differences in ACO quality and could serve as a disincentive for ACOs to engage with such factors. ACOs should and do influence a broad range of patient- and community-level factors that can mitigate the risk of readmission associated with the contextual environment.

We did, however, conduct analyses of SES factors to further inform the Committee's deliberation (see 2b4.4b). To examine the influence of community-level contextual factors, we utilize a patient-level variable, the AHRQ SES index, that is validated as a measure of community-level contextual factors. We also examined the influence of dual Medicare and Medicaid eligibility status on All-Cause Hospital Readmissions (ACR) measure performance.

Examples That Illustrate Empirically Testing Risk Factors

Developers may consider examining the between-entity and within-entity variation, specifically for social and/or functional risk adjustment. A between-entity effect can be described as a scenario in which accountable entities caring for a disproportionate number of patients with social and/or functional risk-vulnerable patients provide lower quality of care to all patient populations compared with accountable entities serving fewer patients with social and/or functional risk. Within-entity effects would account for a scenario in which accountable entities have poorer quality of care for patients with social and/or functional risk compared with patients without social and/or functional risk within the same entity.[80] One approach to estimating these effects is a decomposition analysis, illustrated by example 4 below.

Developers may consider examining the impact of social and/or functional risk factors on the distribution of measured (accountable) entity performance, especially on the lower end of the distribution of performance. However, developers should use caution not to compare measure score performance with clinical risk adjustment, but only to measure score performance with clinical and social or functional risk adjustment in terms of correlations of measure scores or change in rankings or distributions. It is unlikely that a single social or functional factor will make a meaningful difference in the distribution of measure scores or accountable-entity rankings.[80] Example 5 below displays the correlation between a measure outcome with and without social risk adjustment.

Developers may consider examining the thresholds defined in how the measure will be used or implemented. For example, if the measure will be used in an application that defines cutoff for categories of performance (e.g., assigning stars[81] or a payment penalty threshold), developers should examine how social and functional risk factor adjustments influence performance in the context of these thresholds.

Below, Examples 4 and 5 show verbatim text submitted by a measure developer.

Example 4. NQF #2880 Excess Days in Acute Care (EDAC) After Hospitalization for Heart Failure (HF) – NQF-Endorsed (Yale CORE / Centers for Medicare & Medicaid Services)

Statistical Methods:

We assessed the relationship between the social risk factor (SRF) variables with the outcome and examined the incremental effect in a multivariable model. For this measure, we also examined the extent to which the addition of any one of these variables improved model performance or changed hospital results.

One concern with including SRFs in a model is that their effect may be at either the patient or hospital level. For example, low SES may increase the risk of EDAC because patients of low SES have a higher individual risk (patient-level effect) or because patients of low SES are more often admitted to hospitals with higher overall EDAC (hospital-level effect). Identifying the relative contribution of the hospital level is important in considering whether a factor should be included in risk adjustment; if an effect is primarily a hospital-level effect, adjusting for it is equivalent to adjusting for differences in hospital quality. Thus, as an additional step, we assessed whether there was a "contextual effect" at the hospital

level. To do this, we performed a decomposition analysis to assess the independent effects of the SRF variables at the patient and hospital levels. If, for example, the elevated risk of EDAC for patients of low SES were largely due to lower-quality/higher-EDAC risk in hospitals with more patients of low SES, then a significant hospital-level effect would be expected with little-to-no patient-level effect. However, if the increased EDAC risk were solely related to higher risk for patients of low SES regardless of hospital effect, then a significant patient-level effect would be expected, and a significant hospital-level effect would not be expected.

Specifically, for the two selected SRFs (i.e., low SES and dual eligibility), we decomposed the effect of a given SRF on the risk of EDAC as follows: Let Xij denote a binary indicator of the SRF's status of patient *i* at hospital *j* and *Xj* denote the percent of patients with the SRF at hospital *j*. Next, we added *Xij* into the original model adjusting for comorbidities only and broke down Xij = (Xij - Xj) + Xj, in which we let the first component, (Xij - Xj), represent the patient-level social risk variable and the second component, Xj, represent the hospital-level social risk variable. By adding the SRF into the original risk adjustment model and decomposing it into patient- and hospital-level variables, we can simultaneously estimate the SRF's within-hospital or patient-level effect (Xpatient) and between-hospital-level effect (Xhospital) on the risk of EDAC; then, we can assess, after controlling for the effects of comorbidities, whether the two levels of effects are independent and whether one level of effect contributes more than the other. The decomposition analysis allows us to calculate the effects of these two classifying groups: (1) hospitals with higher or lower proportions of low-SES patients or patients dually eligible for Medicare and Medicaid on the risk of EDAC for an average patient and (2) patients' low SES or dual eligibility on their risk of EDAC when they are seen at an average hospital.

It is very important to note, however, that even in the presence of a significant patient-level effect and absence of a significant hospital-level effect, the increased risk could be partly or entirely due to the quality of care patients receive in the hospital. For example, biased or differential care provided within a hospital to low-income patients compared with high-income patients would exert its impact at the level of individual patients and would therefore be a patient-level effect.

It is also important to note that the patient-level and hospital-level coefficients cannot be quantitatively compared because the patient's SES circumstance in the model is binary, whereas the hospital's proportion of low SES patients is continuous. Therefore, in order to quantitatively compare the relative size of the patient and hospital effects, we calculated a range of predicted probabilities of EDAC based on the fitted model.

Specifically, to estimate the average hospital-level effect of an SRF, we calculated the predicted probabilities of EDAC for the following scenarios: (1) assuming all patients did not have the SRF (Xij = 0 for all i and j) and were seen at hospitals with a percent of patients with the SRF at the 5th percentile (P5) of the observed percent of patients with the SRF of all hospitals; (2) assuming all patients did not have the SRF and were seen at hospitals with a percent of patients with the SRF at the 95th percentile (P95); (3) assuming all patients did have the SRF (Xij =1 for all i and j) and were seen at hospitals with a percent of patients with the SRF at the 5th percentile (P5); (4) assuming all patients did have the SRF and were seen at hospitals with a percent of patients with the SRF at the 95th percentile (P95). The

estimated average hospital-level effect is calculated as ((2)-(1) + (4)-(3))/2 (denoted as P95-P5). Then, to estimate the average patient-level effect of an SRF, we calculated the predicted probabilities of EDAC for scenarios, assuming all patients did or did not have the SRF (Xij =0 or 1 for all i and j) and were seen at hospitals with the percent of patients with the SRF at nine selected percentiles (0th, 5th, 10th, 25th, 50th, 75th, 90th, 95th, and 100th). Then, we calculated the difference in predicted probabilities between patients with and without the risk factor who were seen at hospitals with the same percent of patients with the SRF at each of the nine percentiles (DELTAp, p=1, …, 9). We calculated the average of those differences in predicted probabilities as (DELTA1+…DELTA9)/9 (denoted as Delta) as the patient-level effect.

In summary, the difference in predicted probabilities of EDAC for an average patient seen at hospitals with a percent of patients with the SRF at the 95th and 5th percentiles (P95-P5) of hospital percent of patients with the SRF estimates the hospital-level effect of the SRF on the risk of EDAC. We used the 5th and 95th percentiles rather than the maximum and minimum to avoid outlier values. The difference in predicted probabilities between patients with or without the SRF seen at an average hospital (Delta) estimates the patient-level effect of the SRF on the risk of EDAC. If P95-P5 is greater than Delta, it suggests that the hospital-level effect of the SRF is greater than the patient-level effect. That is, the hospital-level effect of the SRF contributes more than the patient-level effect on patients' risk of EDAC.

We also performed the same analysis for several clinical risk variables selected from the comorbidities included in the original risk adjustment model to contrast the relative contributions of patient- and hospital-level effects of clinical risk variables to the relative contributions of the within- and between-hospital level effects of SRFs on patients' risk of EDAC.

Contextual Effect Analysis:

As described, we performed a decomposition analysis for each SRF variable to assess whether there was a corresponding contextual effect. To better interpret the magnitude of results, we performed the same analysis for selected clinical risk factors. The results are described in the tables/figures below.

Most of the patient-level and hospital-level effects of the dual-eligible and low AHRQ SES variables were significant in the logistic and Poisson part of the HF EDAC hurdle model (Table 11). This indicates that both the patient- and hospital-level, dual-eligible effects of the SRFs are associated with an increased risk of acute care and expected duration of that care at the patient and hospital levels.

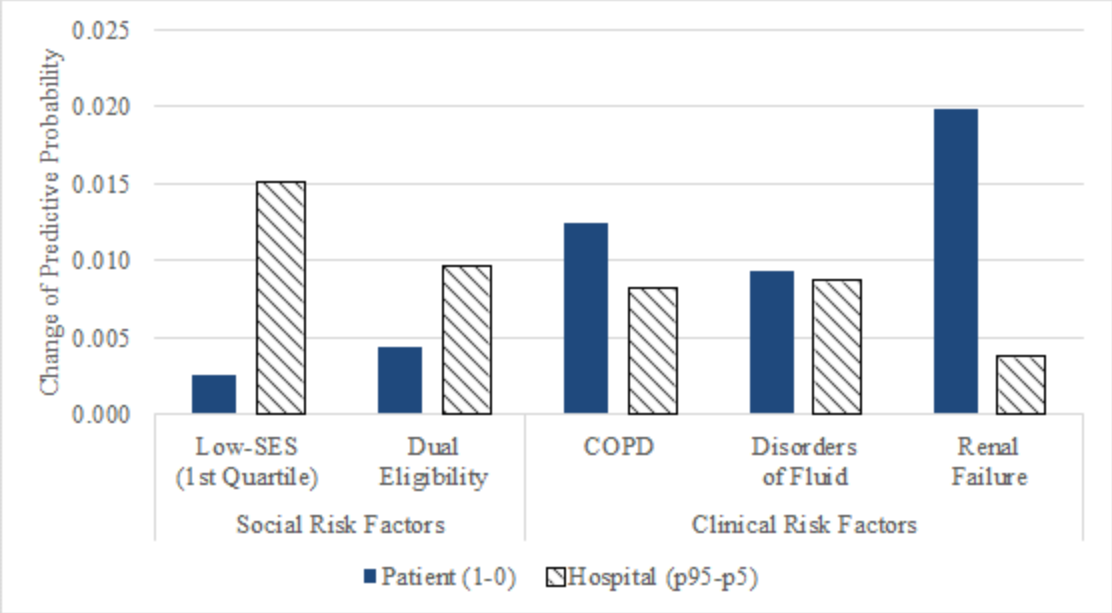
Both the patient- and hospital-level effects contribute to an increased risk; if the dual eligibility and low-SES variables were added into the model to adjust for patient-level differences, then some of the differences in both risk of acute care and expected duration of care between hospitals would also be adjusted for, potentially obscuring a signal of hospital quality.

Table 11. Parameter Estimates for Hospital Level and Patient Level in 2020 From Decomposition Analysis

Parameter	Logistic model Estimate (standard error), p-value	Poisson model Estimate (standard error), p-value
Low AHRQ SES – Patient Level	-0.008 (0.002), p=0.0002	0.047 (0.005), p=<.0001
Low AHRQ SES – Hospital Level	0.068 (0.019), p=0.0003	0.335 (0.018), p=<.0001
Dual-Eligible – Patient Level	-0.001 (0.002) p=0.790	0.060 (0.006), p<.0001
Dual-Eligible – Hospital Level	0.185 (0.025), p<.0001	0.110 (0.025), p<.0001
Chronic Obstructive Pulmonary Disease (COPD) – Patient Level	0.046 (0.002), p<.0001	0.103 (0.004), p<.0001
COPD – Hospital Level	-0.055 (0.032), p=.088	0.659 (0.032), p<.0001
Disorders of Fluid – Patient Level	0.027 (0.002), p<.0001	0.118 (0.005), p<.0001
Disorders of Fluid – Hospital Level	0.576 (0.041), p<.0001	0.003 (0.047), p=0.957
Renal Failure – Patient Level	0.120 (0.002), p<.0001	0.159 (0.005), p<.0001
Renal Failure – Hospital Level	0.527 (0.036), p<.0001	-0.190 (0.041), p<.0001

However, as mentioned above, the patient-level and hospital-level coefficients shown in Table 11 cannot be quantitatively compared because the patient's SES circumstance in the model is binary, whereas the hospital's proportion of low SES patients is continuous. Therefore, to quantitatively compare the relative size of the patient and hospital effects, we calculated a range of predicted probabilities of EDAC based on the fitted model (Figure 4).

Figure 4. Decomposition Analysis Showing the Patient-Level and Hospital-Level Effects for Each Social Risk Factor (HF EDAC)*

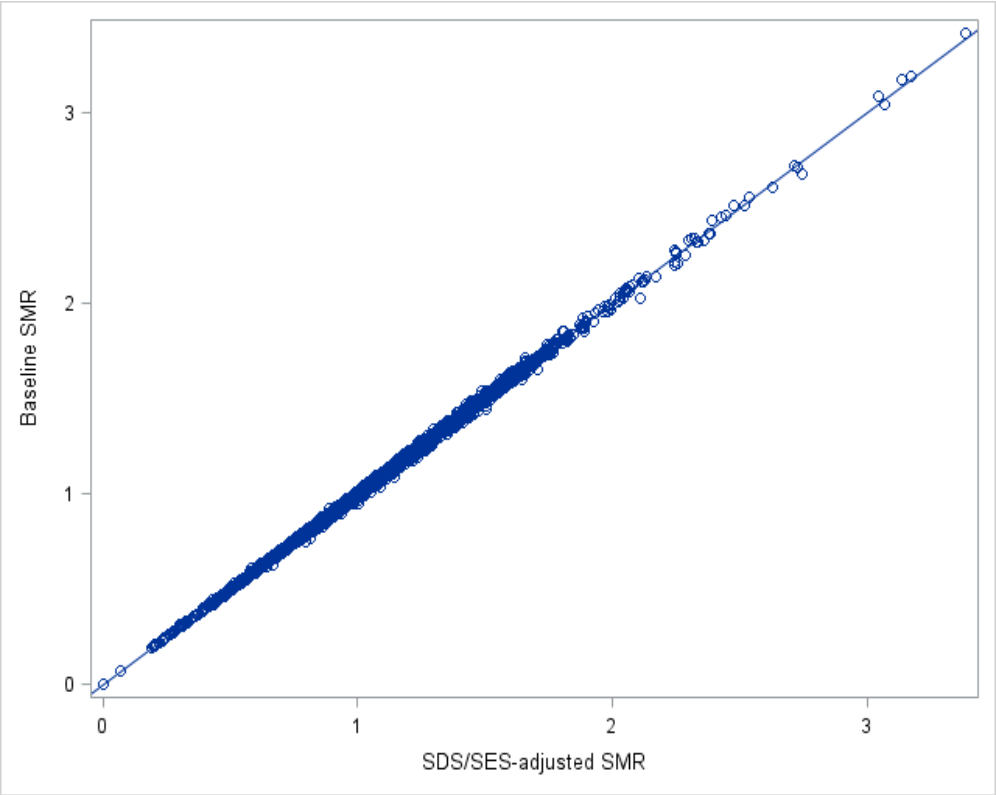


*These values are not comparable to Table 11 because the dual eligibility variable is binary, and the AHRQ SES variable is continuous; therefore, to compare the two, we calculated a range of predicted probabilities of EDAC based on the fitted model.

As shown in Figure 4, as expected, the clinical risk factors shown for comparison have a larger patient-level effect compared with their hospital-level effects. In contrast, both the low AHRQ SES variable and the dual-eligible variable have a larger hospital-level effect compared with the patient-level effect.

Example 5. NQF #0369 Standardized Mortality Ratio for Dialysis Facilities – NQF-Endorsed (University of Michigan Kidney Epidemiology and Cost Center [UMKECC] / Centers for Medicare & Medicaid Services)

Figure 1. Correlation Between Standardized Mortality Ratio (SMR) With and Without SES Adjustment, 2015-2018



ρ = 0.99959

Table 6. Flagging Rates by Model With and Without SES Adjustors: 2015-2018

SHR With SES	Baseline SMR Better Than Expected	Baseline SMR As Expected	Baseline SMR Worse Than Expected	Total
Better Than Expected	129	6	-	135 (2%)
As Expected	4	6,579	5	6,588(95%)
Worse Than Expected	-	5	240	245 (4%)
Total	133 (2%)	6,590 (95%)	245 (4%)	6,969 (95%)

Blank cells marked with (-).

Interpretation:

After adjustment for SDS/SES, 20 facilities (0.29 percent) changed performance categories. Eleven (0.16 percent) facilities were upgraded, and nine (0.13 percent) were downgraded.

Patient race, Hispanic ethnicity, and female sex were associated with lower mortality; however, the impact of these social risk factors is conditional on their respective relationships with other risk factors

captured in the interaction terms in the standardized mortality ratio (SMR). Among SES factors, only unemployment was associated with mortality (higher risk). Neither dual-eligible status nor area-level SES deprivation was associated with mortality. Furthermore, SMRs with and without adjustment for patient SES and area SES are highly correlated and adjustment for SES shifts facility performance only slightly. This suggests that SES does not contribute much to the flagging profiles for facility performance.

Patient-level SES factors are not included in the final risk-adjusted model. In the absence of definitive evidence demonstrating that socioeconomic risk adjustment does not result in differential access to care, the most appropriate decision is not to risk-adjust for socioeconomic factors. While other studies have shown the association between these patient- and area-level SES factors and mortality, further work is needed to demonstrate that differences based on these factors are not related to facility care in order to prevent disparities in care. The primary goal should be to implement quality measures that result in the highest quality of patient care and equitable access for all patients to that care.

In the final SMR model, we continue to include race, ethnicity, and sex for risk adjustment based on results from the literature as discussed in section 2b3.3b. Specifically, the direction of the relationship between race, ethnicity, and mortality is inverted relative to the general population, with lower observed mortality in Blacks and Hispanics on chronic dialysis compared to Whites and non-Hispanics (Kalbfleisch et al 2015). As noted by Kalbfleisch et al, the intent of the measure is to clearly identify facilities whose outcomes are below the national average. With this approach, the adjusted analyses that include race, Hispanic ethnicity, and sex do not obscure disparities in healthcare but tend to clarify potential disparities. Without adjustment, we may erroneously conclude that those facilities with a high concentration of these generally underserved populations have outcomes better than the national norm. Females in the general population have lower mortality rates (Centers for Disease Control and Prevention [CDC] National Vital Statistics Reports, 2012) than males. Adjustment for sex allows for a fair comparison between dialysis facilities with patient populations that have a different mix of males and females.

Examples That Illustrate Empirically Testing the Adequacy of the Risk Model

Below are two examples of testing calibration across lower to higher risk patients:

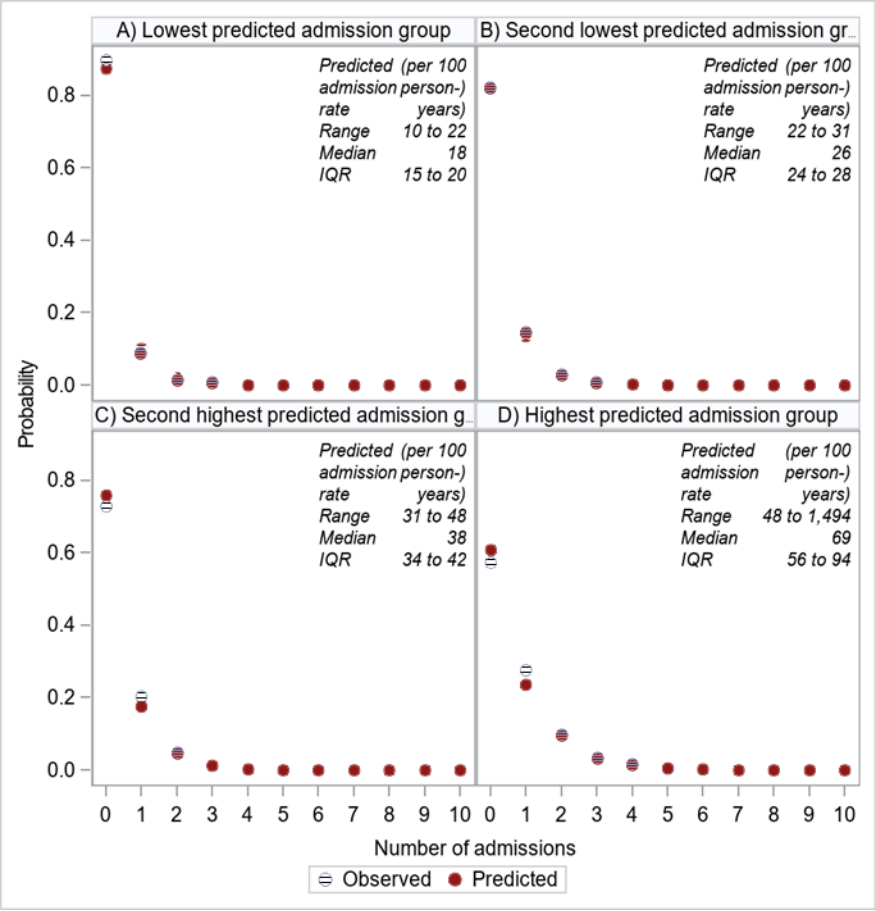
- Example 6 includes risk decile plots for the purpose of examining risk model calibration across subgroups. The plots display observed versus predicted probabilities for a range of outcomes across quartiles of high to low risk.
- Example 7 uses similar methods for a measure of resource use. Both are examples of how previously NQF-endorsed measures have examined model performance in a range of groups of varying levels of risk.

Below, Examples 6 and 7 show verbatim text submitted by measure developers.

Example 6. NQF #3597 Clinician-Group Risk-Standardized Acute Hospital Admission Rate for Patients With Multiple Chronic Conditions Under the MIPS – NQF-Endorsed (Yale CORE / Centers for Medicare & Medicaid Services)

Statistical Risk Model Calibration – Risk Decile Plots or Calibration Curves

A comparison of observed versus predicted probability for the number of hospital admissions among patients with multiple chronic conditions by risk quartile in the 2018 ICD-10 Testing Data Set is shown below.

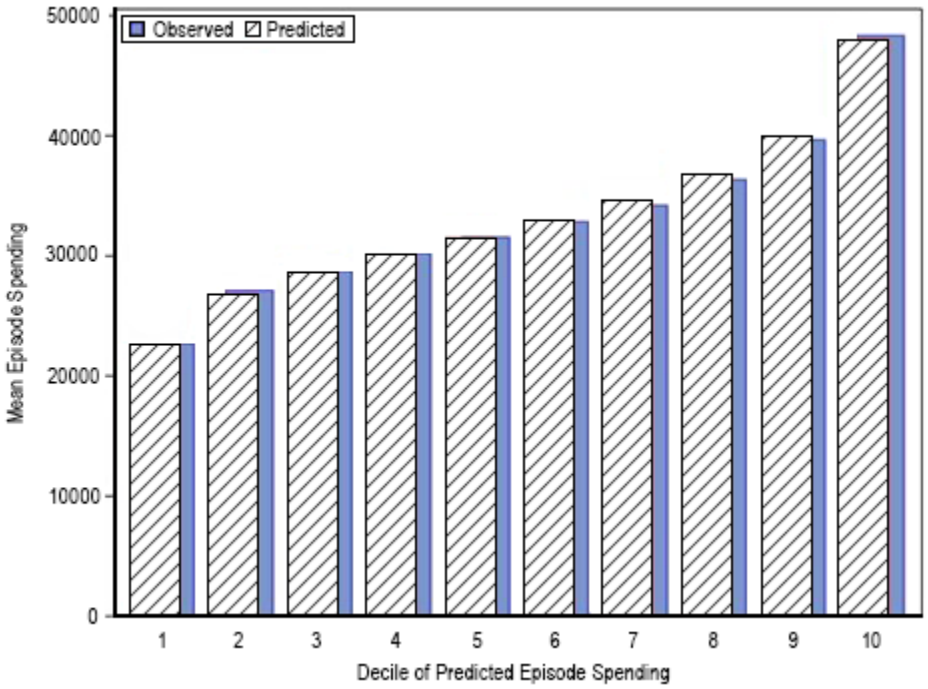


The plots of observed and predicted probabilities for each number of hospital admissions (i.e., 0, 1, 2, …, 10) across quartiles of risk showed that the model performs well across a broad range of risk. In the highest-risk group, we found that the observed and predicted probabilities for zero and one admission differed slightly. However, these differences were small and somewhat expected among the highest-risk group of patients.

Example 7. NQF #3561 Medicare Spending per Beneficiary Post-Acute Care Measure for Inpatient Rehabilitation Facilities – NQF-Endorsed (Acumen / Centers for Medicare & Medicaid Services)

To test the adequacy of this model, we conducted risk-decile testing and plots: We calculated the distribution of episode spending by decile to examine the model's ability to predict both very low- and high-cost episodes. Specifically, we created a "risk score" for each episode calculated as the predicted cost values from each episode divided by the national average of predicted cost value. After arranging episodes into deciles based on the risk score, we calculated the difference and ratio between predicted and observed cost for each decile.

Figure 12. Inpatient Rehabilitation Facilities (IRF) Model Diagnostics: Comparison of Observed and Predicted Spending by Predicted Spending Deciles



Analysis of Medicare Claims File for IRF FY 2016-2017

Table 111. IRF Model Diagnostics: Comparison of Observed and Predicted Spending by Predicted Spending Deciles

Deciles of predicted episode cost	Number of episodes	Observed episode cost	Predicted episode cost	Predicted minus observed cost	Observed / predicted costs
1	61,800	22,702	22,616	-85.61	1.00
2	61,799	27,152	26,783	-368.48	1.01
3	61,799	28,757	28,652	-104.68	1.00
4	61,801	30,242	30,131	-111.18	1.00
5	61,798	31,553	31,490	-63.53	1.00
6	61,799	32,851	32,961	110.31	1.00
7	61,800	34,219	34,629	410.17	0.99
8	61,799	36,357	36,744	386.35	0.99
9	61,799	39,667	39,860	193.02	1.00
10	61,799	48,355	47,989	-366.21	1.01

Analysis of Medicare Claims File for IRF FY 2016–2017.

The model discrimination and calibration results demonstrate good predictive ability across the full range of episodes, from low- to high-spending risk. There was no evidence of excessive under- or overestimation at the extremes of episode risk. The overall adjusted R-squared value is 0.1595. The model controls for over 100 comorbidities (including comorbid interactions), case-mix categories, and patient risk factors. Extensive clinical review was performed by clinicians with experience in providing care in inpatient rehabilitation facilities (IRF) settings in collaboration with medical officers at CMS to identify and review relevant risk factors. Furthermore, certain features of the model improve its policy and practical usability while potentially reducing its fit statistics (i.e., adjusted R-squared value). Most importantly, unrelated services, such as planned hospital admissions and routine management of certain pre-existing chronic conditions (see **section S.9.1 of the Intent to Submit form**), were purposefully and carefully excluded to improve the ability to interpret and compare Medicare Spending per Beneficiary–Post-Acute Care (MSPB–PAC) IRF scores across providers. The R-squared value cannot be evaluated alone and must be considered in combination with the costs excluded from the measure to ensure clinical validity. Since unrelated services may be well predicted by patient risk factors, excluding them can reduce the explained portion of the cost variance and the model's adjusted R-squared value. For example, MSPB–PAC IRF excluded services, such as routine dialysis for end-stage renal disease (ESRD), because they were not believed to be prescribed by or within the scope of the IRF providers. If these services had been included in the IRF measure, doing so would have increased the R-squared value because the ESRD indicator variable in the risk adjustment model would explain much of the

variation due to dialysis. This, however, would have created an inferior measure, as it would lack clinical validity.

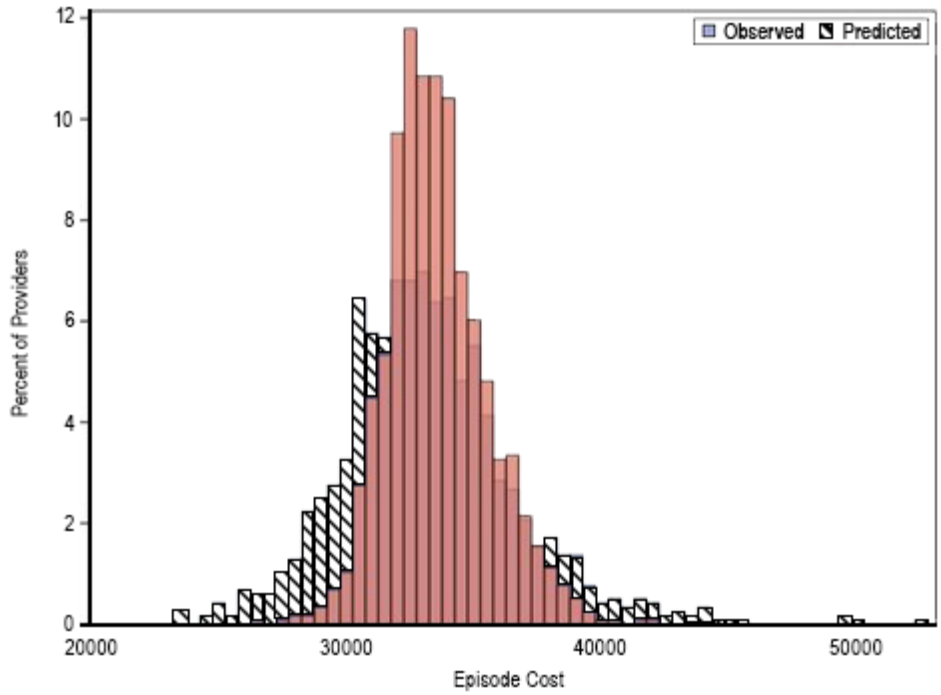
The distribution of facility-level observed and risk-adjusted spending is shown in Table 12Table 122 and Figure 2Figure 23. By considering beneficiary characteristics that are outside of the provider's control, the model compresses the distribution of provider-level spending and decreases its variability. The degree of compression demonstrates that a significant amount of variation in IRF spending exists that is not explained by the observed beneficiary risk factors.

Table 122. Distribution of Provider-Level Observed and Risk-Adjusted Episode Spending

Group	K	Mean	SD	10th Pct	25th Pct	50th Pct	75th Pct	90th Pct
Observed	1,161	33,185.0	3,454.9	29,256.2	31,022.0	32,936.3	34,931.9	37,389.5
Predicted	1,161	33,562.4	1,959.6	31,305.5	32,253.9	33,345.3	34,687.3	36,272.9

Analysis of Medicare Claims File for IRF FY 2016–2017.

Figure 23. Distribution of Provider-Level Observed and Risk-Adjusted Episode Spending



Analysis of Medicare Claims File for IRF FY 2016–2017.

Appendix E: Public Comments

Prior to finalizing documents, National Quality Forum (NQF) solicits comments for a defined period of time via an online tool located on the project webpage. For the *Best Practices for Developing and Testing Risk Adjustment Models* project, the commenting period opened on August 31, 2022, and closed on September 21, 2022. NQF received nine comments from four organizations pertaining to the **draft Technical Guidance**.

Comments received were shared **via memo** with the TEP prior to the public meeting on October 24, 2022. **During the meeting,** the proposed responses and adjudications were presented for the TEP's consideration and discussion.

Appendix F: Stakeholder Feedback Memo

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** 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. Furthermore, NQF and the TEP developed the Technical Guidance during 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 risk adjustment 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 given 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 the 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., the 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 **Table 2**.

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.

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 the **Focus Group Topics** section. 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, given 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 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 considered as part of the 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 subgroups, several focus group participants noted that without further specificity with respect to identifying the relevant subgroups, developers may spend a significant amount of time determining multiple relevant subgroups, 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 subgroup(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 problems with small number reliability. 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 whether 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 currently do not consistently stratify results by the quality of care provided to patients who are from the LGBTQI+ 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 subgroup if the intent is to simply display differences in performance for that subgroup. However, if a certain subgroup is at a higher risk for a

measured outcome, then stratifying the measure separately for this subgroup 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 six data categories for race and ethnicity, 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 that 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/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 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
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. • 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 an 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 guidance 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

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.*

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. 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.

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.

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 needed that accountable entities can mitigate the impact of the social or functional risk factors on 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 subgroups (e.g., racial categories). Moreover, these subgroups should be defined in the conceptual model.

Additionally, all risk models should be tested and vetted to examine whether they 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[27] 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.