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 – VERSION 4
August 30, 2021

This Technical Guidance document 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. In this context, the overall performance measure score is used to make a conclusion about the quality of a healthcare entity (i.e., a hospital, health plan, practice, or other entity that is being assessed) in relation to other entities or some other comparator, such as average performance. Such comparisons should be affected as little as possible by factors other than quality of care, such as patient characteristics already present at the start of care.

Because healthcare outcomes are a function of patient attributes (e.g., clinical, social, and functional factors) as well as the care received, and since healthcare entities do not have the same mix of patients, risk adjustment is essential to ensuring an "apples-to-apples" comparison when examining outcome performance in real-world settings. Risk adjustment (also known as case-mix adjustment) refers to statistical methods to control or account for patient-related factors when computing performance measure scores. Risk-adjusting outcome and cost/resource use performance 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 is widely accepted. With the increased use of these measures within public reporting and payment programs comes increased scrutiny of the adequacy and fairness of the risk adjustment methodologies used, especially as it relates to social risk factors and functional status-related risk factors. Additionally, there is an increased focus on leveraging quality measures to promote health equity. 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. However, approaches to the risk adjustment of these factors vary, ranging from the data sources and statistical models used to the steps taken to determine whether these factors are included in the overall risk model. As a result, measure developers, stewards, and program implementers have expressed a need for standardization and guidance in developing, testing, and evaluating risk adjustment models that account for social and/or functional risk.

Through input from an NQF-convened Technical Expert Panel (TEP), this Technical Guidance document describes a step-by-step approach to developing and testing risk adjustment models that account for social and/or functional status-related risk factors within quality measurement. The intent of this guidance is to provide measure developers with a standard approach to social and/or functional risk adjustment within performance measurement. Furthermore, this guidance identifies best practices, as minimum standards, for risk adjustment models. These minimum 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]).

NQF recognizes that each performance measure must be assessed individually to determine the appropriateness of social and/or functional status-related risk adjustment. Beginning with the conceptualization stage, it is important to illustrate the concepts of social and/or functional risk that have an impact on the modeled system, care pathway, framework, etc. The conceptual model will set the foundation for determining the types of factors to consider within the model and whether to risk-adjust, to stratify, to do both, or neither. This guide further explores the testing methodologies that developers may consider for statistically analyzing risk factors for inclusion in the model and for the overall adequacy of the model. Lastly, as the field of quality measurement changes rapidly, this guidance will need to continue to evolve to align with the advancements in quality measurement science.

Introduction

Background

Over the last decade, the quality measurement enterprise has rapidly moved towards linking payment to quality of care, generally known as value-based purchasing (VBP), to improve health delivery and health system accountability. For VBP to be successful, 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. In addition, accountable entities need comprehensive, reliable, 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 level the playing field, **risk adjustment** methods have been applied to many quality performance measures, but not all, and not in a standardized manner across measures.[1]

Risk-adjusting outcome and cost/resource use performance 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.[2,3] However, the increased use of outcome and 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[4–6]), and **social risk factors,** such as income, education, social support, neighborhood deprivation, and rurality.[7,8] Functional risk factors are important to examine since they may confound the relationship between social risk, quality outcomes, and resource use.

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 as the coronavirus 2019 (COVID-19) pandemic continues to unfold.[9–11] 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. The pandemic underscores the importance of exploring and appropriately adjusting for all applicable social risk factors to ensure accurate assessment and to prevent inappropriate financial penalization of accountable entities due to caring for patient populations with increased social and/or functional risk.[12]

NQF recognizes that **health equity** is fundamental to all quality improvement efforts. Quality measurement should contribute to closing the health equity gap and not inadvertently institutionalize it. 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. 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 a key component for successful health outcomes across the nation. As social risks are increasingly recognized to have a tremendous impact on health and healthcare outcomes, NQF recognizes that fully addressing inequities associated with race/ethnicity and

social risks requires a holistic policy approach and a private-public sector partnership that goes well beyond the purview of quality measurement. There is a clear distinction between directly adjusting payment rates with social risk factors and adjusting quality measures that may be tied to financial bonuses and incentives. This report only focuses on the latter case—whether and how to adjust quality measures for social risk factors so that accountable entities will be compared fairly. Quality measure adjustment alone cannot and should not be used to achieve resource (re)allocations.

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. Alternatively, the inclusion of social and functional risk factors in risk adjustment models may not make transparent the differences in care outcomes. To mitigate the latter concern, this guidance instructs developers to stratify measure results by key risk factors. **Risk stratification** is an important tool to deploy in conjunction with risk adjustment to identify healthcare disparities. 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).[12–16]

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.[13] In 2014, NQF convened a Risk Adjustment Expert Panel, which recommended allowing risk adjustment when there is a conceptual rationale and empirical relationship present.[3] The NQF Board of Directors implemented a trial period in 2015, during which adjusting measures for social risk factors was no longer prohibited.[17] 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 validity of a performance measure in truly reflecting care quality and resource use.[18] These efforts have demonstrated that social risk adjustment may be feasible and appropriate, but it remains challenging for many measure developers to obtain 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 that account for social and/or functional risk. Approaches to risk adjustment of these factors requires consideration of the data sources and statistical models used, the specific risk factors used to represent functional status, social determinants of health (SDOH), socioeconomic status (SES), sociodemographic status (SDS), and how to determine whether these factors should be included in the overall risk model. Hence, developing a standardized, consistent approach to risk adjustment would facilitate accurate assessment of the role of functional, social, and clinical risks; enable fair, unbiased comparisons of performance of the accountable entities with different patient case mix; and report and monitor disparities across subpopulations.[18]

Purpose

This Technical Guidance document 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. Although it is uncommon, there may also be a relationship between social/functional risk factors and some process measure scores (e.g., filling a drug prescription could be affected by patient's SES as in **NQF #0541**, which is adjusted for age, gender, low-income subsidy (LIS)/dual status, and disability status).[19,20] Certain measures, such as serious reportable events (SREs) or never events, should not be risk-adjusted since they are largely preventable and indicative of a problem in a healthcare setting's safety systems. Instead, social and functional risk factors should be stratified for reporting.

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 good and emerging best practices, as minimum requirements, for social and/or functional risk adjustment within performance measure development.

The intent of this Technical Guidance document 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 cost/resource use measures should be adjusted for social and functional risk within VBP programs.

Project Overview

With a goal of advancing measurement science, NQF developed this Technical Guidance document for measure developers; it includes good and emerging best practices, as minimum requirements (referred to hereafter as *minimum standards*), for social and/or functional risk factor adjustment in quality performance measure development. To accomplish this goal, NQF, with support from the Centers for Medicare & Medicaid Services (CMS), convened a multistakeholder TEP (**Appendix A**) in the fall of 2020 to provide input and guidance on the current state of risk adjustment for social and functional status in measurement, emerging good and/or 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.

During the first phase of this effort, 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 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 frameworks.

Results of the environmental scan were used to facilitate the development of the Technical Guidance document. Together with the input and diverse perspectives shared by the TEP, this guidance describes the process of conceptualizing an outcome or a cost/resource use performance measure and the

subsequent risk adjustment model development (specifically accounting for social and/or functional risk) and decision making that will be needed for NQF endorsement review.

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.

Healthcare disparities refer to the differences between groups in health insurance coverage, access to and use of care, and quality of healthcare services.[21]

Health disparities refer to a higher burden of illness, injury, disability, or mortality experienced by one group relative to another.[21]

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.[21,22,23]

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.[24] Functional status encompasses both the individual's ability to carry out ADLs and to participate in life situations and society.[25] 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 mediators). Functional status encompasses the whole person and is affected by physical, developmental, behavioral, emotional, social, and environmental conditions.[24]

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.[5]

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.[1]

Social risk factors are the social conditions or factors that may have a conceptual and empirical relationship to healthcare outcomes.[26] Illustratively, these factors may include socioeconomic position/status (e.g., income, education, and occupation), race/ethnicity/linguistic and cultural context, gender, social relationships, residential and community environments, urbanicity/rurality, and health literacy. Additionally, this guidance includes a variety of socioeconomic and demographic factors as social risk factors (e.g., age, Medicare/Medicaid dual eligibility, and uninsured). For this guidance, age is treated as both a clinical and social risk factor.[26]

Social determinants of health (SDOH) are the social, nonmedical conditions that determine healthcare provision and health outcomes.[26] They can both improve and worsen an individual's health.

Social or functional status-related risk adjustment refers to statistical adjustment for sociodemographic and/or functional status-related variables.

Stratification refers to an approach to address social or functional risk factors in the performance measurement process. In addition to reporting overall performance, stratification consists of computing performance separately for different strata or groupings of patients based on some characteristic(s) (i.e., each healthcare unit has multiple performance scores, one for each stratum rather than one overall performance score).[13]

Core Principles

To ground this Technical Guidance document 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*.[13,27] 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 the aims of the CMS Quality Management Action Plan[29]
- 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 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/ethnicity variables incorporate elements of social risks, such as environment, access to high quality care, genetically mediated predispositions to certain diseases and/or different responses to treatment (including medications). In situations in which only race and ethnicity data are available but other specific variables (e.g., granular social risk data; detailed, personalized genetic information) are not, the inclusion of variables such as race/ethnicity may be the best available—though imperfect—variables to serve as proxies for social risk factors

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.[28] Commonly used methods include an assessment of 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 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 widely different methods have been used across similar measures, which emphasizes 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 document addresses this need by highlighting good and emerging 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 document serves as a resource for risk adjustment model development and testing that 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 that regard, the guide describes risk adjustment of social and functional risks across five main steps:

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*

To round out and evolve prior NQF guidance,[13] a key new direction in this Technical Guidance document includes an increased emphasis on the conceptual model and a decrease in overly prescriptive empirical testing requirements. Additionally, to align with national efforts to leverage quality measurement to promote health equity[29] and to further mitigate concerns that risk adjustment of social and functional risks will mask disparities, this guidance establishes a minimum standard, which includes requiring measures to be stratified in conjunction with risk adjustment in order to improve the ability to measure health disparities and differential outcomes.

Beginning with a conceptual model, developers are encouraged to consider the big picture, namely how the patient-level clinical, functional, and social risk factors, that are present at the start of care (i.e., measurement period), influence the measured outcome and how the accountable entity can mitigate these factors to lower risk. 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 provider locus of control. Once risk factors have been identified in the conceptual model, the guide moves to the next step, which is to identify and select data sources and variables for inclusion in the model. At this phase of the process, developers should carefully examine the various data quality considerations, including the potential bias that may be introduced due to data availability challenges.

After the appropriate data sources have been identified, the Technical Guidance document reviews testing methodologies for statistically analyzing risk factors for inclusion in the model, followed by the overall adequacy of the model (i.e., calibration and discrimination tests of the risk adjustment model in subpopulations specific to the measure). It should be noted 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 support the conceptual model; some of these have been added as illustrative examples from NQF-endorsed measures (**Appendix D**). This aligns with the core principle of not being overly prescriptive as to limit the use of novel methods or to add significant burden to measure developers.

Lastly, the decision to adjust or not adjust for social and/or functional risk requires not only an empirical assessment of the risk model, but also a consideration of the potential unintended consequences and healthcare policies. 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. However, developers should take a balanced and thorough consideration of the trade-offs in adjusting for social and/or functional risks.

As the field of quality measurement changes rapidly, this guidance will also need to evolve to align with advancements in measurement science. The information collected in this guide reflects the TEP's decisions and recommendations. To that regard, this guidance acknowledges several emerging data sources, drawing attention to the future of quality measurement. 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.

Standard Risk Adjustment Framework

This guidance identifies good and emerging best practices as minimum standards, supporting each of the steps in this process. These standards form a framework for risk adjustment of health outcomes and offer a robust path forward to achieving reliable and valid measure scores 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. 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 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 requirements for social and/or functional risk adjustment. 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. These standards attempt to balance measurement theory with the practical limitations of measure development. It is also not meant to diminish the investigation into diseases and processes that need novel measure development. Rather, these recommendations 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.

Risk Adjustment

1	A conceptual model is required and should illustrate the pathway between the social and/or functional status-related risk factors, patient clinical factors, quality of care, and the measured healthcare outcome
2	Developers should consider age, gender, race/ethnicity, urbanicity/rurality, Medicare and Medicaid dual eligibility, indices of social vulnerability (such as the Area Deprivation Index and Agency for Healthcare Research and Quality [AHRQ] SES Index score) and markers of functional risk (such as frailty, ADLs, and instrumental ADLs [IADLs]) in the conceptual model.
3	If social and/or functional status risk factors are not available but are included in the conceptual model, the developer should describe the potential bias that may exist and the direction and describe the magnitude of that bias as a result of not including the risk factor(s) in the model. The developer should also provide a justification for why the measure still has validity even 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 those variables. Developers should also provide a description of the populations covered within that 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	Calibration should be conducted within the overall population and within relevant at-risk clinical, social, and functional subpopulations. All risk models should be tested and vetted to examine whether they significantly under- or over predict 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 and their interpretation of the results.
7	Risk stratification should be tested in conjunction with risk adjustment to maximize the ability is able to identify healthcare disparities.

NQF-endorsed measures are "**best in class**" because multistakeholder committees achieve consensus agreement on a set of standard **endorsement criteria**. This guidance will further advance NQF's measure evaluation criteria by introducing new minimum standards for social and functional risk adjustment. These minimum standards are listed below:

Conceptualizing the Model

Developing the Conceptual Model

A conceptual model visually describes the pathway between the social and/or functional status-related 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 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.

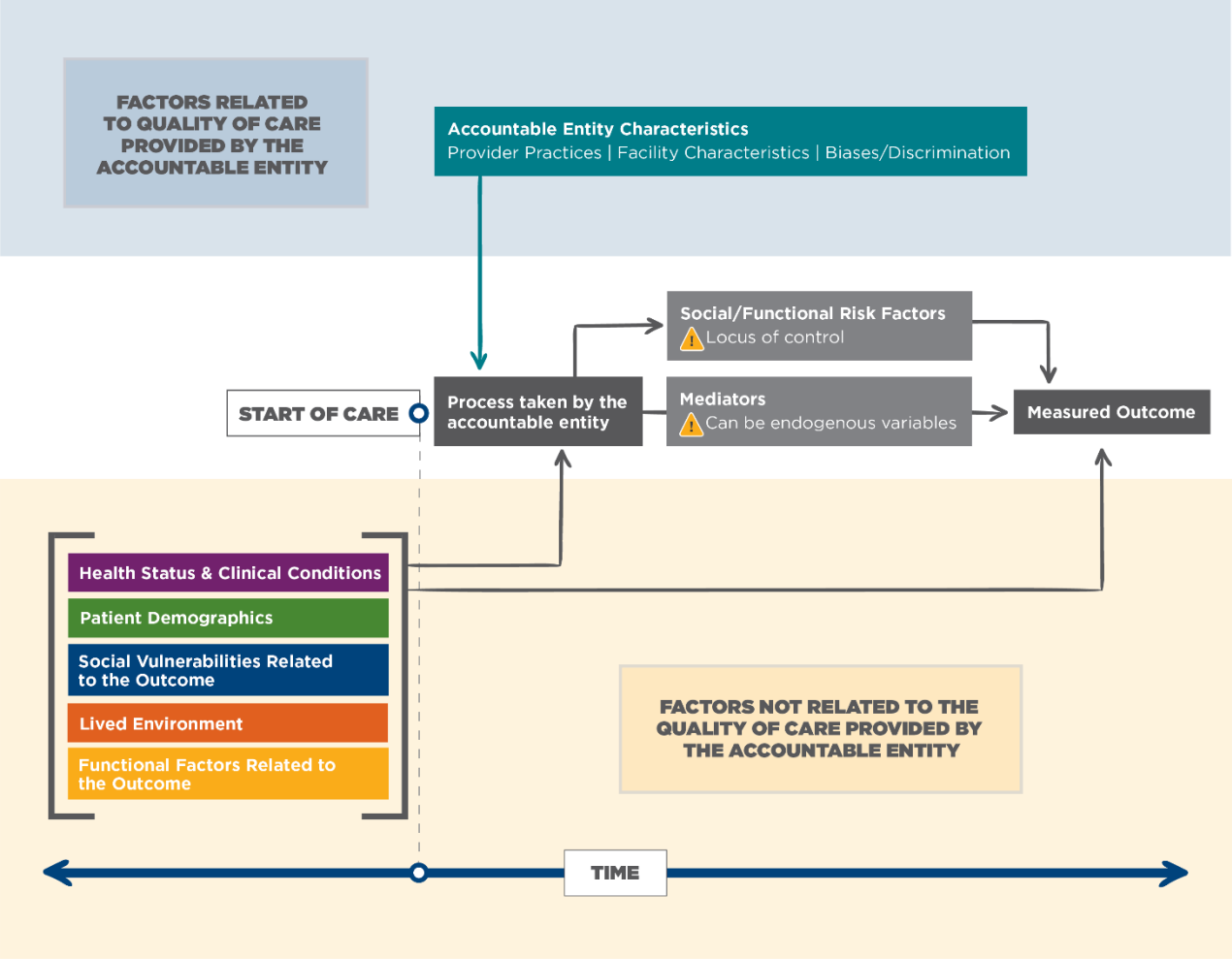
Risk adjustment is based on patient characteristics at the start of care (i.e., measurement period). Therefore, all demographic, clinical risk factors, social and functional risks, and patient preferences related to the outcome of interest should be considered for inclusion in the conceptual model, regardless of whether the data can be operationalized in the full measured population. As described in minimum standard #3 above 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 that will have an impact on their use as variables within the risk model. Nevertheless, this should not prohibit the consideration of these factors within the conceptual model, as these factors should be identified in the conceptual model regardless of whether they can be operationalized in available data.

MINIMUM STANDARD

A conceptual model is required and should illustrate the pathway between the social and/or functional status-related risk factors, patient clinical factors, quality of care, and the measured healthcare outcome.

It is strongly recommended that developers construct a graphical representation of these relationships for clarity and ease of analysis. Below is a graphical depiction of a standard conceptual model (Figure 1). It depicts a template for developers to use to visualize the basic structure of a conceptual model. The remainder of this section then describes considerations for identifying the contents of this template conceptual model. An example graphic from a CDP measure submission is also presented 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 not related to the quality of care provided by accountable entities (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). This is 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 symbolizes the accountable entity's characteristics and can include, but is not limited to, provider practices (e.g., adequate discharge planning), potential biases/discrimination (e.g., inability to provide or prioritize translation services or culturally competent care), and facility characteristics (e.g., safety-net providers or critical access facilities). Because this box contains characteristics related to 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 for adjustment for the measure of interest.

MAPPING THE PROCESS OF CARE AND THE MEASURED OUTCOME

The middle portion of the graphic contains the care pathway. Developers need to map this section so that they can determine which factors should and should not be included in 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 symbolizes 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 measured entity's locus of control. Yellow caution signs denote that developers should think carefully about these boxes in particular, given the subsequent steps in this Technical Guidance document.

Arrows from the patient characteristics to the process of care and the measured outcome boxes illustrate that characteristics already present at the start of care (i.e., start of measurement period) can influence these elements of the care pathway. The arrow from the accountable entity's characteristics to the process of care illustrates that the measured entity's characteristics influence the processes available and/or chosen to achieve the measured outcome.

AREAS OF SPEICAL 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 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 the provider has the ability to meaningfully influence the social and/or functional risk factor, the risk factor would be placed in the top portion of the figure. If not, the risk factor would be placed at the bottom of the figure to make clear its potential for inclusion in the risk adjustment model. This can be done by citing literature or conducting empirical analyses.

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 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 pathway between the actions of the accountable entity and the measured outcome of readmissions (i.e., an 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 risk-adjusted away, considering that they lie along the care pathway and relate to the quality of care of the accountable entity.

REITERATING THE PURPOSE OF THE CONCEPTUAL MODEL FIGURE

The conceptual model serves as the foundation for the remaining steps outlined in this Technical Guidance document. The risk adjustment model can be misleading and ineffective unless it is grounded in a transparent conceptual model informed by the literature and expert input (i.e., clinicians, patients). Developers should also write a brief description of their processes for developing the conceptual model, using citations to establish the relationship between factors and outcomes. This will help others who are uninvolved in the conceptual model's development to understand what decisions were made and why.

The remainder of this section is concerned with the steps for developing the contents of a conceptual model.

Factor Selection for Inclusion

First, measure developers should explore the broad list of factors that might have an impact on the outcome. These factors can be identified by a combination of expert opinions, literature review of peer-reviewed articles and white papers, and previous work on quality measures in the disease or topic area. 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.[30] Measure developers will also look to the 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 advocates) may also be involved in order to verify or further examine the impact these risk factors can have on the 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 be reflected in the conceptual model. They could be eliminated during the testing phase when developers are able to identify any statistical issues (e.g., overfitting, multicollinearity, and/or confounding) in the model's structure to remove these biases from the model. Once the conceptual model is fully drafted, developers should review their results from end to start. Moving backwards through the model can help to identify assumptions that were made or logical fallacies that may otherwise go unnoticed.[31]

When designing the conceptual model, it is important to remember that these factors can have either a direct or indirect (i.e., via the actions taken by the accountable entity) effect on the measured outcome.[32] Both 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 that also contribute to quality of care and quality measurement.

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 best 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 into and exclusion out of the final risk model should be mindfully considered, especially for factors in which there is disagreement on their impact. There are a number of social and functional risk variables that should always be considered in the conceptual model for outcome and cost/resource measures. Based on the environmental scan, the TEP 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 should be considered, at a

minimum, for examination in conceptual models: age, gender, race/ethnicity, urbanicity/rurality, Medicare and Medicaid dual eligibility, indices of social vulnerability (such as the Area Deprivation Index or the Agency for Healthcare Research and Quality [AHRQ] SES Index score), and markers of functional risk (such as frailty, ADLs, and instrumental ADLs [IADLs]). The consideration of these factors within the conceptual model is not a requirement for their use in the final risk adjustment model, as this is dependent upon their relationship to the outcome of interest. However, developers should describe the rationale for determining whether to include or exclude them from the conceptual model.

MINIMUM STANDARD

Developers should consider age, gender, race/ethnicity, urbanicity/rurality, Medicare and Medicaid dual eligibility, indices of social vulnerability (such as the Area Deprivation Index and Agency for Healthcare Research and Quality [AHRQ] SES Index score) and markers of functional risk (such as frailty, ADLs, and instrumental ADLs [IADLs]) in the conceptual model.

Variable Selection for Examination

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. 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 subpopulations within which they will test the calibration of the model, as mentioned **later in this guide**, and make clear in the conceptual model the reasons why subpopulations may be affected by certain risk factors differently.

Within the conceptualization of the model, developers should carefully consider the use of **proxy factors**. Proxies can be introduced when developers identify and select potential data sources and variables. A clear explanation of the relationship between the proxy factor and the unmeasured social or functional risk concept is vital.

Level of Measurement

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 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. For example, considerations might include the degree of control that accountable entities have to influence outcomes, which may vary by context. For example, ACOs can potentially influence certain social risk factors, such as transportation barriers (at a cost), whereas individual clinicians may have less ability to do so. Therefore, the same measure developed for ACOs and individual clinicians may require different risk adjustment models.

Developers should consider whether social and/or functional risk factors confound the quality-outcome relationship. Specifically, what is the level of evidence that accountable entities can mitigate the impact of regarding the social or functional risk factors of 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.

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 and provider locus of control. To the extent known by the developer at initial measure submission, the specific intended use of the measure should be explained. 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. The intended use should be balanced with the locus of control 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.

Measures tied to strong financial incentives that are used for VBP should consider the evidence regarding how accountable entities can take specific actions to mitigate the relationship between social and/or functional risk and the outcome. In VBP scenarios, it is important to reduce the potential for risk aversion, especially for some providers (e.g., safety net) who serve a disproportionate number of patients with social and/or functional risk factors. The conceptual model should outline the evidence in context of the locus of control and specific intended use of the measure. Moreover, developers should re-evaluate social and/or functional risk adjustment when adapting measures for other uses.

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 that will have an impact on 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.

MINIMUM STANDARD

If social and/or functional status risk factors are not available but are included in the conceptual model, the developer should describe the potential bias that may exist and the direction and describe the magnitude of that bias as a result of not including the risk factor(s) in the model. The developer should also provide a justification for why the measure still has validity even in this circumstance.

Transparency is one of the core principles of risk adjustment. Developers must ensure these data are reliable, valid, complete, comprehensive, timely, and generalizable (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., are the data periodically audited?). 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+ Medicare).

MINIMUM STANDARD

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 (Table 1). Developers can cite other research to show data quality of those variables. Developers should also provide a description of the populations covered within that 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; and 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.

Consideration	Description
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 **Empirically Testing the Adequacy of the Risk Model**).

Different statistical rules apply to different types of models. For example, a model with an outcome that is more common may require as few as 30 cases per accountable entity to consistently return the same model statistics across samples. If the outcome is uncommon, then the number of cases required could be much larger.[1] 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). A statistician can provide guidance to determine the appropriate sample sizes based on the characteristics of the sample(s) and the requirements of the types of analyses in use. In general, the larger the sample size is, the greater the statistical power to detect outliers and the higher the measure reliability will be.

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 electronic health records (EHRs). Of these, the most common data source for developing risk adjustment models is claims data, namely Medicare Fee-for-Service (FFS) claims.

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.[33] Similarly, the CMS-sponsored Post-Acute Care Interoperability (PACIO) project is developing item sets

for cognitive impairment and frailty, areas of functional status that have had ambiguous definitions and scarce data.[34]

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.[35] However, Z codes are currently not widely used in claims. Developers should exercise caution with the use of Z codes within risk adjustment models due to their limited availability. 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.[36] 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[37], the ACS Data[38], the Area Deprivation Index[39], or the Bayesian Indirect Surname Geocoding[40] 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 consider empirically testing 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. Empirically screening social and/or functional risk factors is not deterministic in measures with large sample sizes since these screening tests for admission to the adjustment model create barriers based on an implicit assumption that introducing an additional variable substantially increases variance. 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 accurate adjustment for fair comparison, rather than predicting or clearly distinguishing 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.[41,42] The rationale to exclude certain social and functional 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 to 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 subpopulations of interest. **Appendix D** provides several illustrative examples of empirical testing approaches that developers may consider. Although the empirical testing is not deterministic, developers should examine that evidence in conjunction with the conceptual model. Developers should also describe the statistical methods used and the results and interpretation of the analyses, all of which leads to the decision of whether or not to select social and/or functional risk factors for risk adjustment. Developers should be transparent about their approach and their interpretation of the results.

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 (accountable) entities. Absolute or relative frequency statistics are examples of descriptive statistics that can be used for discrete social and/or functional risk factors.[43] 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. It is not intended to make inferences or judgements on whether the factor is appropriate for inclusion in the risk adjustment model. However, variables with little or no variation in frequency across measured entities are not likely to be of value in modeling performance differences across accountable entities, even if these factors have a significant association with outcomes.

MINIMUM STANDARD

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

Empirically Testing the Adequacy of the Risk Model

Measure developers should assess the risk adjustment model to ensure that it does not violate important underlying assumptions (e.g., distributional). The ability to assess model performance is subject to the same data limitations identified when selecting data sources for risk model variables. However, measure developers should assess the model to determine its predictive ability, discriminant ability, and overall fit.

In order 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 model adequacy. Measure submissions should provide statistical results from testing the approach to control for differences in patient characteristics.

There are various approaches to assessing the performance of a risk adjustment model. One approach is using measures such as explained variation (e.g., R^2 statistics) to quantify how close expected predictions are to the observed outcome. Risk model discrimination is a critical step in identifying whether patients who have the observed outcome have a higher expected risk than those with a lower risk expectation. This can be quantified with measures of sensitivity, specificity, or area under the receiver operating characteristic curve (AUC or c-statistic).[44]

When considering the contribution of social risk and/or functional risk factors in modeling decisions, developers may compare the discrimination performance, such as AUC for risk adjustment models that include social and/or functional risk factors and models that include clinical factors only. However, improvement in the AUC may not always recognize important social and/or functional risk factors in terms of an increase in the AUC, especially if the standard, clinical-factor-only model has a large baseline AUC.[45] Changes in model discrimination, such as c-statistics, may not be sufficient to inform a decision on whether to include an additional social and/or functional risk factor in the model specification.[46] Another useful 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).[45,47,48]

Risk adjustment model performance must also 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. 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). As a result, graphical approaches may be preferred (e.g., plots of observed-to-expected outcomes across a broad range of expected values). To adequately assess the impact of social and/or functional risk, risk adjustment model calibration must be examined within at-risk subpopulations (e.g., racial categories). These subpopulations should be defined in the conceptual model.

MINIMUM STANDARD

Calibration should be conducted within the overall population and within relevant at-risk clinical, social, and functional subpopulations. 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 and their interpretation of the results.

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 reasons for considering a separate model for certain population subgroups. For example, if there are data suggesting that one SES/racial group has considerably different outcomes than another or that the association of other covariates with outcomes is considerably different for that subgroup, then it could be argued that they would be better served by their own model. Calibration and other performance metrics would probably be better, but it would be at the cost of losing generalizability of the model to other populations. Another reason for a separate model would depend on the policy goal 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. Lastly, developers should use caution when building separate models on subgroups unless there is sufficient sample size.

Considerations for Determining the Final Risk Adjustment Model

Social and/or functional risk adjustment may not be appropriate for all measures. 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. 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.[49]

Additionally, when performance measures are used for accountability applications, such as public reporting and pay-for-performance, then purchasers, policymakers, and other users of performance measures 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 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, a balanced and thorough consideration and discussion of the tradeoffs in adjusting for social and/or functional risk is a critical element of this standardized framework.

Risk Stratification

Risk stratification refers to the division of a population or resource services into distinct, independent strata or groups of similar data, thus enabling 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 differences in results. Risk stratification is an important analysis to conduct in conjunction with risk adjustment to identify health and healthcare disparities.

MINIMUM STANDARD

Risk stratification should be tested in conjunction with risk adjustment to maximize the measure's ability to identify healthcare disparities.

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.[50] Developers should report stratification specifications (e.g., categories and combinations of

social risk factors) by specific and relevant subgroup categories, such as racial/ethnic categories, gender, SES, and functional status.[27] 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. This stratification should also align with the intended use of the measure, if known. For instance, if a CMS **Quality Improvement Program** stratifies quality measure results by race, Medicare/Medicaid dual-eligible status, disability status, LGBTQ+, and SES, then the developer should provide this information for NQF endorsement review.

Risk stratification further supports quality improvement program gap evaluation decisions, such as those conducted by NQF's Measure Applications Partnership (MAP). Since 2011, MAP has been convened by NQF and funded by CMS to recommend high quality performance measures that address national healthcare priorities, fill critical measurement gaps, and increase alignment of measures among public and private measurement programs. This gap evaluation, facilitated by risk stratified results, can promote health equity in care and the elimination of healthcare disparities.[51]

Negative Unintended Consequences

Historically, risk adjustment of quality performance measures has focused primarily on clinical factors (e.g., pre-existing medical conditions). Whether or not to incorporate social risk and/or functional status risk factors, however, continues to be ardently debated due to concerns that it could have negative unintended consequences. Most importantly, there is a concern that adjusting for social risk may harm patients with social risk factors by not making potential differences in performance transparent. Adjusting potentially weakens the ability of the measure to drive improvements in care for specific at-risk groups by setting lower standards for the very populations who most need care improvements. However, without appropriate risk adjustment, VBP programs may create perverse incentives, such as underdelivering otherwise appropriate and beneficial care for patients with social and/or functional risk factors because of their anticipated worse unadjusted outcomes.[26]

Adjusting for social and/or functional risk factors may benefit patients with these risk factors. Unadjusted measures may lead to inappropriate financial penalties among providers who care for patients with a high proportion of social risk and who are unable to mitigate the subsequent increased risk of the measured outcome. These financial penalties may leave some providers who care for disadvantaged populations with fewer resources for quality improvement activities.[52,53] Using other mechanisms to assist such providers may be useful, such as additional training or financial resources for those caring for more at-risk populations. Due to the differential in resources that may be required to achieve a measured outcome in an at-risk population, compensation 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.[54]

Regardless, continuous monitoring for potential unintended, adverse consequences is needed to mitigate any impact on patients with social and/or functional risk factors. 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).

Policy Considerations

In its recent report to Congress, the Department of Health and Human Services' (HHS) ASPE concluded that resource use measures used in VBP programs should be adjusted for social risks, whereas many outcome measures should not.[14] 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.

NQF Policy

NQF takes into consideration the guidance in the ASPE report. Similar to the ASPE report, this Technical Guidance document acknowledges the importance of data collection and interoperability standards for social and functional risk. This Technical Guidance document also supports the need to improve risk adjustment overall to meet the demands of a changing healthcare landscape, which includes the creation of a standard risk adjustment framework that includes social and functional risk for risk-adjusted outcome and resource use measures.[16] To that regard, this Technical Guidance document describes a framework of minimum standards that developers should consider for social and/or functional risk adjustment within quality measurement. Similar to ASPE's policy statement, NQF also recognizes that while measurement approaches are critical, simultaneous payment approaches that award and support better outcomes for vulnerable persons need to be developed to fully account for social risks.[55]

However, unlike ASPE's recommendations in its second report to Congress, this guidance does recommend that quality and resource use measures be adjusted for social and/or functional risk factors based on the conceptual model. Current NQF endorsement criteria are agnostic to measure use, including use within VBP arrangements, which is the focus of ASPE's recommendations. This TEP and other NQF-convened groups, such as the **NQF Scientific Methods Panel** (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 dependent on the intended use (i.e., evaluating validity and reliability for each use type). While this 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, this TEP recommends that developers re-evaluate social and/or functional risk adjustment when adapting measures for other uses.

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.

Quality measures are first and foremost a tool to improve care for patients by drawing inferences on accountable-entity performance on the measured outcome. This relationship and the locus of control of the accountable entity should be identified in the conceptual model. Yet even if performance measures are adjusted for social and/or functional risk factors, this does not ensure protection of certain accountable entities, such as safety net providers. Therefore, additional strategies may be needed.[13] For example, social risk factor adjustment or stratification for 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 comparative performance results. Given that safety net providers are differentially funded (i.e., a function of local and state taxing jurisdictions), making comparisons even among safety net providers may be problematic. Accountability programs should consider whether and how to incorporate this type of community factor into comparative evaluations for purposes of assigning rewards and penalties.[56] 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.[14,55] 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 Merit-Based Incentive Payment System (MIPS) program.

Conclusion

As the U.S. continues to move towards 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 lower-quality care being delivered 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 convened a TEP to provide input on technical guidance for measure developers, which includes emerging best practices on when and how to adjust for functional and social risk factors in measure development. The TEP identified several minimum standards that are rooted in core principles of quality measurement and risk adjustment science. 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 TEP emphasized the importance of first establishing a sound 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 factors 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.

A Path Forward

This Technical Guidance document serves as a resource for both novice and experienced measure developers to develop risk adjustment models that account for social and functional risk factors within outcome and cost/resource use performance measures. The intent of this guidance is to further support NQF-endorsement considerations, in which there has been a perceived need for clarity in the evaluation of these risk models. This guide will facilitate consistency in the evaluation of these risk models through a set of minimum standards 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. NQF will continue to seek to advance measurement science in this important area by engaging relevant stakeholders to garner feedback on the feasibility and utility of this guidance. This feedback will be instrumental in updating the guidance and subsequent NQF measure evaluation criteria and policies to reflect the ever-changing healthcare landscape.

References

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

2 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.

3 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.

4 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.

5 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/.

6 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.

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

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

9 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.

10 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.

11 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.

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

13 National Quality Forum (NQF). NQF: Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors. Accessed March 2021 at https://www.qualityforum.org/Publications/2014/08/Risk_Adjustment_for_Socioeconomic_Status_or_Other_Sociodemographic_Factors.aspx. Last accessed March 2021.

14 Office of the Assistant Secretary for Planning and Evaluation (ASPE). *Office of the Assistant Secretary for Planning and Evaluation Report to Congress 2020, "Social Risk Factors and Performance in Medicare's Value-Based Purchasing Program*.

15 Guinan M. NAM Releases Final Report on Social Risk Factors and Medicare Payment. America's Essential Hospitals. Accessed June 2021 at https://essentialhospitals.org/quality/nam-releases-final-report-on-social-risk-factors-and-medicare-payment/. Published January 13, 2017.

16 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.

17 National Quality Forum (NQF). NQF: Adjusting Measures for Social Risk Factors. Accessed January 2021 at http://www.qualityforum.org/Adjusting_Measures_for_Social_Risk_Factors.aspx#:~:text=NQF%3A%20Adjusting%20Measures%20for%20Social%20Risk%20Factors&text=The%20National%20Quality%20Forum's%20(NQF,a%20person's%20level%20of%20education.

18 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.

19 National Quality Forum (NQF). NQF: Primary Care and Chronic Illness Final Report - Fall 2019 Cycle. https://www.qualityforum.org/Publications/2021/02/Primary_Care_and_Chronic_Illness_Final_Report_-_Fall_2019_Cycle.aspx. Last accessed June 2021.

20 Centers for Medicare & Medicaid Services. Medicare 2021 Part C & D Star Ratings Technical Notes. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwiqzNGmn5XyAhXQMlkFHR-GAg4QFjAAegQIBRAD&url=https%3A%2F%2Fwww.cms.gov%2Ffiles%2Fdocument%2F2021technotes20201001.pdf-0&usg=AOvVaw3tnnDnxLVBcq-0qSZ_P18e.

21 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-5-key-question-and-answers/. Published May 11, 2021. Last accessed June 2021.

22 Robert Wood Johnson Foundation (RWJF). What is Health Equity? What is Health Equity? Accessed June 2021 at https://www.rwjf.org/en/library/research/2017/05/what-is-health-equity-.html. Published May 1, 2017. Last accessed June 2021.

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

24 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.

25 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.

26 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; 2016:21858. http://www.nap.edu/catalog/21858. Last accessed January 2021.

27 National Quality Forum (NQF). NQF: A Roadmap for Promoting Health Equity and Eliminating Disparities: The Four I's for Health Equity. Accessed May 2021 at https://www.qualityforum.org/Publications/2017/09/A_Roadmap_for_Promoting_Health_Equity_and_Eliminating_Disparities__The_Four_I_s_for_Health_Equity.aspx. Last accessed May 2021.

28 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.

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

30 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.

31 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.

32 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/.

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

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

35 American Hospital Association. ICD-10-CM Coding for Social Determinants of Health. Accessed March 2021 at https://www.aha.org/system/files/2018-04/value-initiative-icd-10-code-social-determinants-of-health.pdf.

36 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.

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

38 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.

39 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/.

40 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.

41 Marc Elliott, Amelia Haviland. Use of a web-based convenience sample to supplement a probability sample. *Surv Methodol*. 2008;33(2):211-215.

42 Cefalu M, Elliott MN, Hays RD. Adjustment of Patient Experience Surveys for How People Respond. *Med Care*. 2021;59(3):202-205.

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

44 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.

45 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.

46 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.

47 Pencina MJ, D'Agostino RB, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med*. 2011;30(1):11-21.

48 Pencina M, D'Agostino R, Demler OV. Novel metrics for evaluating improvement in discrimination: net reclassification and integrated discrimination improvement for normal variables and nested models. *Stat Med*. 2012;31(2):101-113.

49 Vogel W, Chen GJ. An introduction to the why and how of risk adjustment. *Biostat Epidemiol*. 2020;4(1):84-97.

50 National Quality Forum. Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement. https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=88439. Published September 2019.

51 Centers for Medicare and Medicaid Services, Center for Clinical Standards and Quality. *Program-Specific Measure Needs and Priorities.*; 2020. https://www.cms.gov/files/document/cms-measurement-priorities-and-needs.pdf.

52 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.

53 McCarthy C, Vaduganathan M, Patel K, et al. Association of the New Peer Group–Stratified Method With the Reclassification of Penalty Status in the Hospital Readmission Reduction Program. *JAMA Netw Open*. 2019;2(4):e192987.

54 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.

55 Steven Sheingold, Rachael Zuckerman, Nancy DeLew, et al. Health Equity And Value-Based Payment Systems: Moving Beyond Social Risk Adjustment | Health Affairs Blog. https://www.healthaffairs.org/do/10.1377/hblog20210726.546811/full/. Last accessed August 2021.

56 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.

57 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/.

58 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.

59 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.

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

61 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.

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

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

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

65 DeCarlo M. Scientific Inquiry in Social Work Glossary. In: *Scientific Inquiry in Social Work*. Open Social Work Education; 2018. Accessed May 2021 at https://viva.pressbooks.pub/scientificinquiryinsocialwork/back-matter/glossary/. Last accessed June 2021.

66 Hidalgo B, Goodman M. Multivariate or Multivariable Regression? *Am J Public Health*. 2013;103(1):39-40.

67 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.

68 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, and NQF Staff

Committee 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

Patrick Campbell, PhD, PharmD, RPh
Associate Director, Policy Research, Merck
Formerly Senior Director, Measurement Outcomes Research, Pharmacy Quality Alliance
Alexandria, Virginia

Elizabeth Drye, MD, SM
Senior Director, Quality Measurement Programs**,** Yale-New Haven Hospital, Center for Outcomes Research and Evaluation (CORE)
New Haven, Connecticut

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

Bellinda King-Kallimanis, PhD, MSc
Director of Patient-Focused Research, LUNGevity
Bethesda, Maryland

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

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

Katherine Vickery, MD, MSc
Clinician-Investigator, Hennepin Healthcare
Minneapolis, Minnesota

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)

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

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)

NQF Staff

Kathleen Giblin, RN
Acting Senior Vice President, Quality Measurement

Kim Ibarra, MBA, MS
Senior Managing Director, Quality Innovation

Matthew K. Pickering, PharmD
Senior Director, Quality Measurement

Taroon Amin, PhD, MPH
Consultant

Sai Ma, PhD
Consultant

Monika Harvey, MBA, PMP
Project Manager, Quality Measurement

Janaki Panchal, MSPH
Manager, Quality Measurement

Hannah Ingber, MPH
Senior Analyst, Quality Measurement

Juanita Rogers, MS, CHES
Analyst, Quality Measurement

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.[57]

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).[58,59]

Collinearity refers to the relationship between two variables when one is highly linearly correlated with the other.[60]

Confounders refer to variables that are related to both the intervention and the measured outcome.[1]

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.[61]

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

Healthcare disparities refer to differences between groups in health insurance coverage, access to and use of care, and quality of healthcare services.[21]

Health disparities refer to a higher burden of illness, injury, disability, or mortality experienced by one group relative to another.[21]

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.[21,22,23]

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.[62] Functional status covers both the individual carrying out ADLs and the individual participating in life situations and

society.[25] 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.[24]

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*.[63]

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.[64]

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.[65,66] 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.[67]

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.[1,59]

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.[5]

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*.[1]

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.[1]

Social risk factors are the social conditions or factors that may have a conceptual and empirical relationship to healthcare outcomes.[26] Illustratively, these factors may include socioeconomic position/status (e.g., income, education, and occupation), race/ethnicity/linguistic and cultural context, gender, social relationships, residential and community environments, urbanicity/rurality, and health literacy. Additionally, this guidance includes a variety of socioeconomic and demographic factors as social risk factors (e.g., age, Medicare/Medicaid dual eligibility, and uninsured). For this guidance, age is treated as both a clinical and social risk factor.

Social determinants of health (SDOH) are the social, nonmedical conditions that determine healthcare provision and health outcomes.[26] They can both improve and worsen an individual's health.

Social or functional status-related risk adjustment refers to statistical adjustment for sociodemographic and/or functional status-related variables.

Stratification (or risk stratification) refers to an approach to address social or functional risk factors in the quality measurement process. In addition to reporting overall performance, stratification consists of computing performance separately for different strata or groupings of patients based on some characteristic(s) (i.e., each healthcare unit has multiple performance scores, one for each stratum rather than one overall performance score).[13]

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.[1]

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.[58,59]

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 in regard to receiving the submissions in a timely manner May be limited to specific demographics, such as 65+ Medicare beneficiaries
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.

Data Source	Strengths	Limitations
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.
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 underenrollment 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

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

Conceptualizing the Model

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) 30-Day Heart Failure (HF) Readmission measure (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 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 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]:

Socioeconomic position

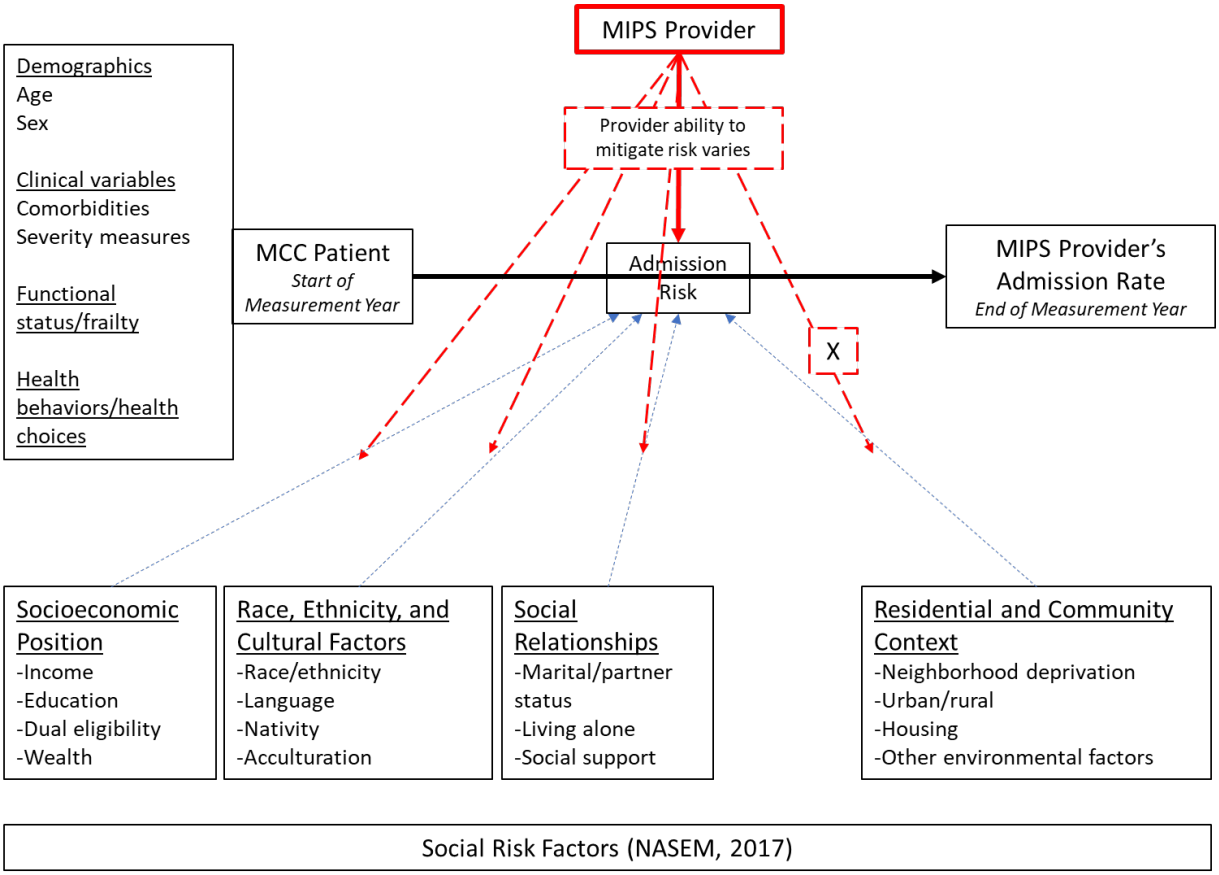
Race, ethnicity, and cultural factors

Social relationships

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 of these factors to 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.

[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.

Variable Selection Guided by the Conceptual Model

NQF #1789 Hospital-Wide Readmission Measure (HWR) – NQF-Endorsed (RTI International / 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 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 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 and sixty-nine articles were initially reviewed, and 155 studies were excluded from 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 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 (ACO):

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.

Empirically Testing in a Multivariable Model

Developers may consider examining the contribution of the social and/or functional risk factors using multivariable modeling. A multivariable analysis helps to understand the relationship of social and/or functional risk factors in relation to the other variables in the model and the outcome(s) being measured simultaneously. Common testing methods include logistic regression and other multivariable analyses. Developers should use caution in interpreting a lack of statistical significance of social and/or functional variables in multivariable models, as an individual social and/or functional factor is unlikely to have a high magnitude of significance due to the number of risk factors in the model that may mediate the relationship.[58] To the extent that social and/or functional risk factors are independent of quality and unmodifiable by the measured (accountable) entity, social and/or functional risk adjustment should generally be included in the risk adjustment model.

#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)

Prevalence of each risk variable and the associated rate ratios for variables in the final risk model

MIPS MCC Cohort

n = 4,659,922

Variable	Prevalence of risk factors n (%)	Adjusted rate ratio (95% CI)
Crude rate (per 100 person-years)	39.1	
Total number of admissions	1,608,763	
Total person time at risk (in years)	4,110,499	
Demographic		
Age <70 y/o	740,962 (15.9%)	
Age 70 to <75 y/o	1,033,292 (22.2%)	1.09 (1.08, 1.10)
Age 75 to <80 y/o	966,205 (20.7%)	1.24 (1.23, 1.25)
Age 80 to <85 y/o	823,759 (17.7%)	1.44 (1.43, 1.45)
Age >=85 y/o	1,095,704 (23.5%)	1.78 (1.77, 1.80)
Nine chronic disease groups		
AMI	100,719 (2.2%)	1.09 (1.08, 1.10)
ALZHEIMERS AND RELATED DISORDERS	1,279,891 (27.5%)	1.27 (1.26, 1.27)
ATRIAL FIBRILLATION	1,167,393 (25.1%)	1.17 (1.17, 1.17)
CHRONIC KIDNEY DISEASE	2,383,858 (51.2%)	1.22 (1.21, 1.22)
COPD/ASTHMA	1,613,996 (34.6%)	1.22 (1.21, 1.22)
DEPRESSION	1,685,967 (36.2%)	1.07 (1.06, 1.07)
HEART FAILURE	1,823,667 (39.1%)	1.36 (1.36, 1.37)
STROKE/TRANSIENT ISCHEMIC ATTACK	635,160 (13.6%)	1.09 (1.08, 1.09)
DIABETES	2,717,638 (58.3%)	1.10 (1.10, 1.10)
Clinical comorbidities Defined using Condition Categories (CCs) or International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes	*	*

Variable	Prevalence of risk factors n (%)	Adjusted rate ratio (95% CI)
Dialysis status (CC 134)	89,380 (1.9%)	1.54 (1.52, 1.55)
Respiratory failure (CC 82, 83, 84)	459,865 (9.9%)	1.13 (1.12, 1.13)
Liver disease (CC 27 [remove K767], 28, 29, 30)	111,999 (2.4%)	1.23 (1.22, 1.24)
Pneumonia (CC 114, 115, 116)	714,580 (15.3%)	1.19 (1.18, 1.19)
Septicemia/shock (CC 2)	314,053 (6.7%)	1.05 (1.04, 1.06)
Marked disability/frailty (CC 21, 70, 71, 73, 157, 158, 159, 160, 161, 189, 190)	569,620 (12.2%)	1.23 (1.23, 1.24)
Hematologic/al diseases (CC 46 [remove D593], 48)	501,562 (10.8%)	1.03 (1.02, 1.03)
Advanced cancer (CC 8, 9, 10, 13)	263,183 (5.6%)	1.21 (1.20, 1.22)
Infectious and immune disorders (CC 1, 3, 4, 5 [remove A1811], 6, 47, 90)	261,668 (5.6%)	1.07 (1.06, 1.08)
Severe cognitive impairment (CC 50 [remove F05, F061, F068], 64, 65, 80)	370,777 (8.0%)	1.09 (1.09, 1.10)
Major organ transplant status (CC 132, 186)	39,216 (0.8%)	1.09 (1.08, 1.11)
Pulmonary heart disease (ICD-10-CM I2601, I2602, I2609, I270, I271, I272, I2789, I2781, I279, I280, I281, I288, I289)	197,778 (4.2%)	1.14 (1.14, 1.15)
Cardiomyopathy (ICD-10-CM I420, I421, I422, I425, I426, I427, I428, I429, I43, I514, I515)	397,841 (8.5%)	1.08 (1.08, 1.09)
Gastrointestinal disease (CC 31, 32, 33, 35, 36)	993,104 (21.3%)	1.06 (1.06, 1.07)
Iron deficiency anemia (CC 49)	2,058,339 (44.2%)	1.13 (1.13, 1.14)
Ischemic heart disease except AMI (CC 87, 88, 89, 98; add ICD-10 I511, I512)	2,415,379 (51.8%)	1.15 (1.14, 1.15)
Other lung disorders (CC 112 [remove J470, J471, J479], 118)	1,939,225 (41.6%)	1.02 (1.01, 1.02)
Vascular or circulatory disease (CC 106, 107, 108, 109 [remove I701, I722])	2,220,460 (47.7%)	1.13 (1.13, 1.14)
Other significant endocrine disorders (CC 23 [remove E748, N251, N2581])	278,126 (6.0%)	1.03 (1.03, 1.04)
Other disabilities and paralysis (CC 72, 74, 103, 104, 119)	292,693 (6.3%)	1.08 (1.08, 1.09)
Substance abuse (CC 54, 55, 56)	578,732 (12.4%)	1.21 (1.21, 1.22)
Other neurologic disorders (75, 77, 78, 79, 81, 105)	1,565,850 (33.6%)	1.09 (1.09, 1.10)
Specified arrhythmias and other heart rhythm disorders (CC 96 [remove I480, I481, I482, I4891] and 97)	1,412,343 (30.3%)	1.05 (1.05, 1.05)
Hypertension (CC 95)	4,204,973 (90.2%)	1.06 (1.05, 1.07)
Hip or vertebral fracture (CC 169, 170)	240,679 (5.2%)	1.07 (1.06, 1.08)
Lower-risk cardiovascular disease (CC 91, 92, 93)	1,260,360 (27.0%)	1.03 (1.02, 1.03)
Cerebrovascular disease (CC 102 [remove I6789])	267,201 (5.7%)	1.06 (1.05, 1.06)
Morbid obesity (ICD-10-CM E6601, Z6835, Z6836, Z6837, Z6838, Z6839, Z6841, Z6842, Z6843, Z6844, Z6845)	600,726 (12.9%)	1.04 (1.04, 1.05)

Variable	Prevalence of risk factors n (%)	Adjusted rate ratio (95% CI)
Urinary disorders (CC 142 [remove N131, N132, N1330, N1339, Q620, Q6210, Q6211, Q6212, Q622, Q6231, Q6232, Q6239] and 145 [remove N2589, N259, N261, N269, Q6102, Q612, Q613, Q614, Q615, Q618])	1,370,375 (29.4%)	1.05 (1.04, 1.05)
Psychiatric disorders other than depression (CC 57, 59, 60, 62, 63 [remove F4321])	1,332,385 (28.6%)	1.08 (1.07, 1.08)
Frailty indicators Defined using Noridian Policy Groups for DME or original reason for Medicare entitlement		
Walking aids	231,405 (5.0%)	0.98 (0.98, 0.99)
Wheelchairs	193,552 (4.2%)	1.13 (1.12, 1.14)
Hospital bed	75,885 (1.6%)	1.09 (1.08, 1.10)
Lifts	17,136 (0.4%)	1.03 (1.01, 1.05)
Oxygen	383,219 (8.2%)	1.38 (1.38, 1.39)
Original Reason for entitlement: DIB (may or may not have ESRD)	685,924 (14.7%)	1.25 (1.24, 1.26)
Original Reason for entitlement: ESRD (may or may not have DIB)	19,072 (0.4%)	1.24 (1.21, 1.27)
Social risk factors		
Low AHRQ SES index score (<=25th pct)	847,802 (18.2%)	1.08 (1.07, 1.08)
Low specialist density (<=25th pct)	167,684 (3.6%)	1.04 (1.03, 1.05)

Assessing the Between-Entity Effects Versus Within-Entity Effects

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.[58]

Developers may also consider examining the independent effects of social and/or functional risk factors at the patient level and at the level of the accountable entity using a decomposition analysis.

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 (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.

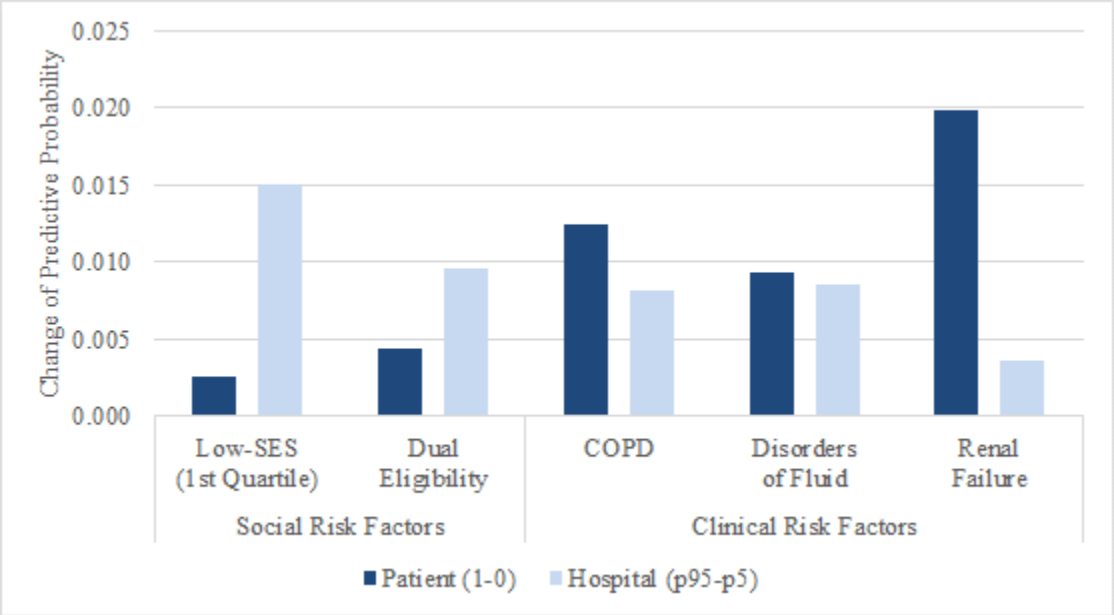
Both the patient- and hospital-level effects contribute to an increased risk; if the dual eligibility and low-SES variables were added into the model to adjust for patient-level differences, then some of the differences in both risk of acute care and expected duration of care between hospitals would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

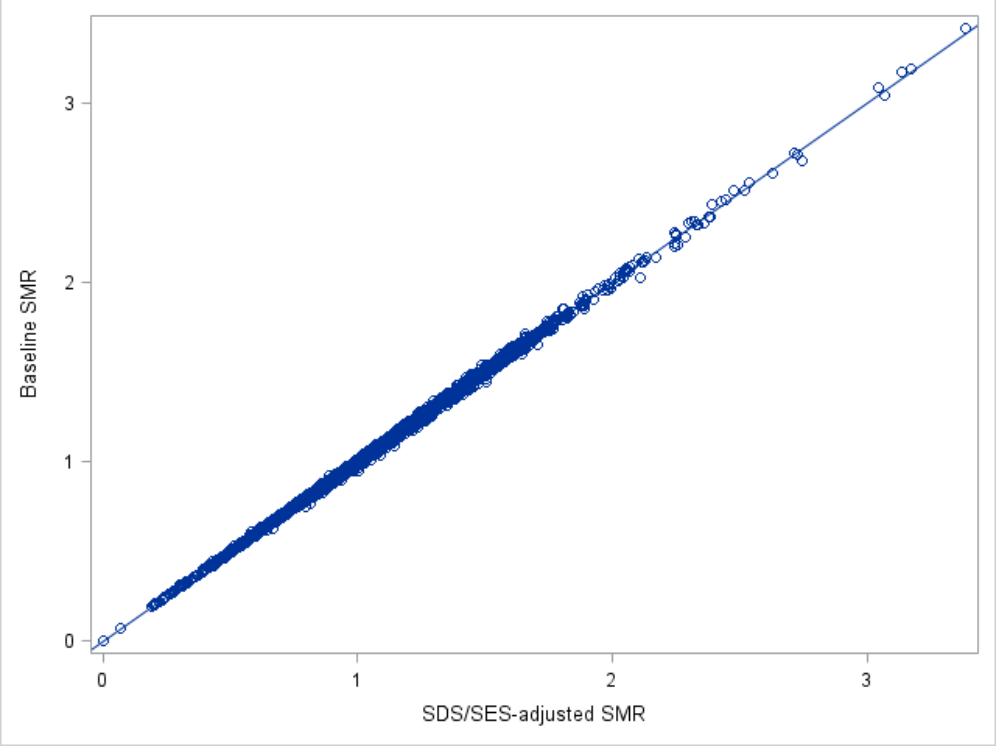
Determining the Impact of Adjusting for Risks (or not) on Accountable Entities in the Tails of the Performance Distribution

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, 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.[58]

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[68] or a payment penalty threshold), developers should examine how social and functional risk factor adjustments influence performance in the context of these thresholds.

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%)

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 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.

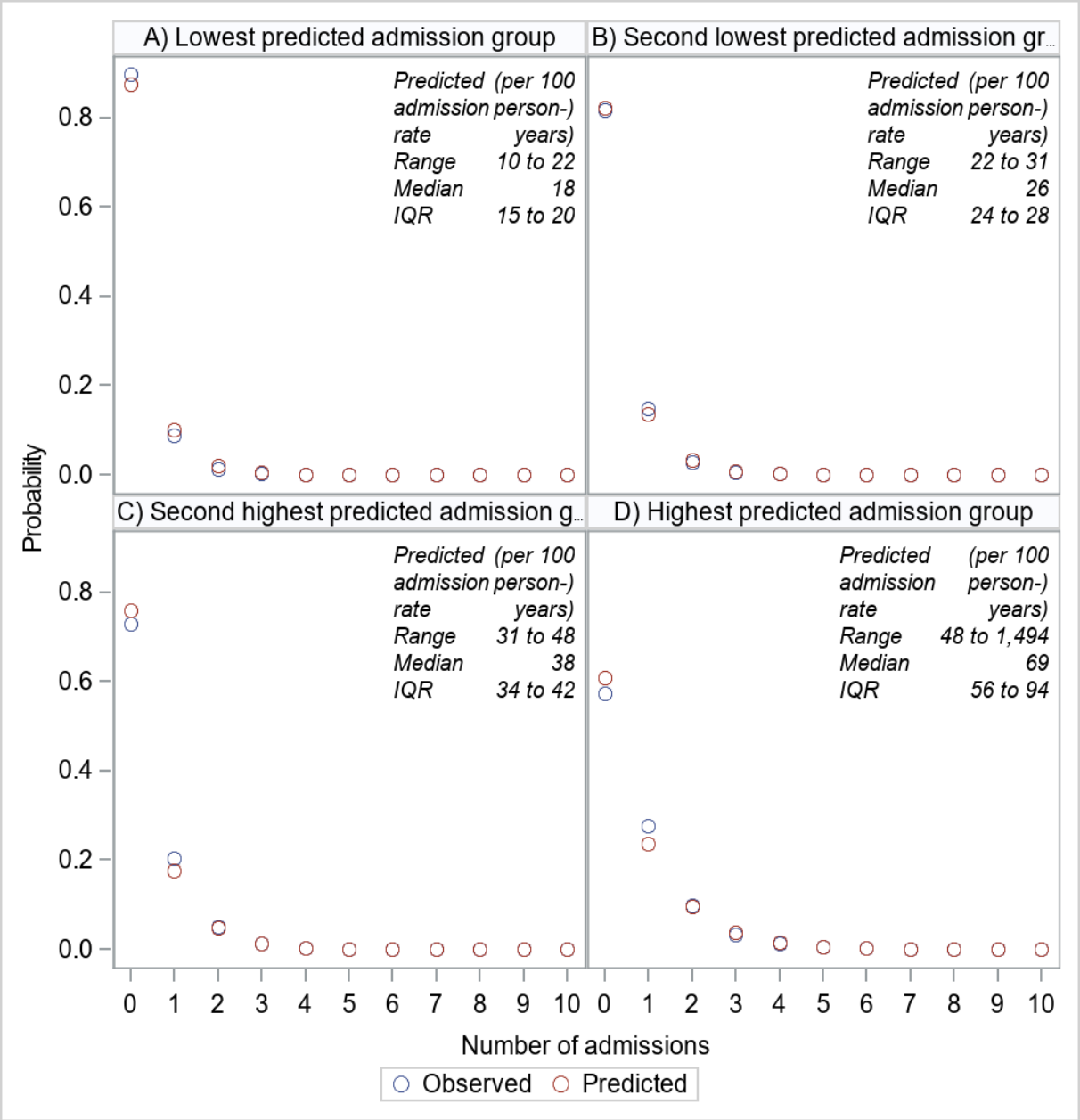
In the final SMR model, we continue to include race, ethnicity, and sex for risk adjustment based on results from the literature as discussed in section 2b3.3b. Specifically, the direction of the relationship between race, ethnicity, and mortality is inverted relative to the general population, with 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.

Risk Model Calibration

Example 1. 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.



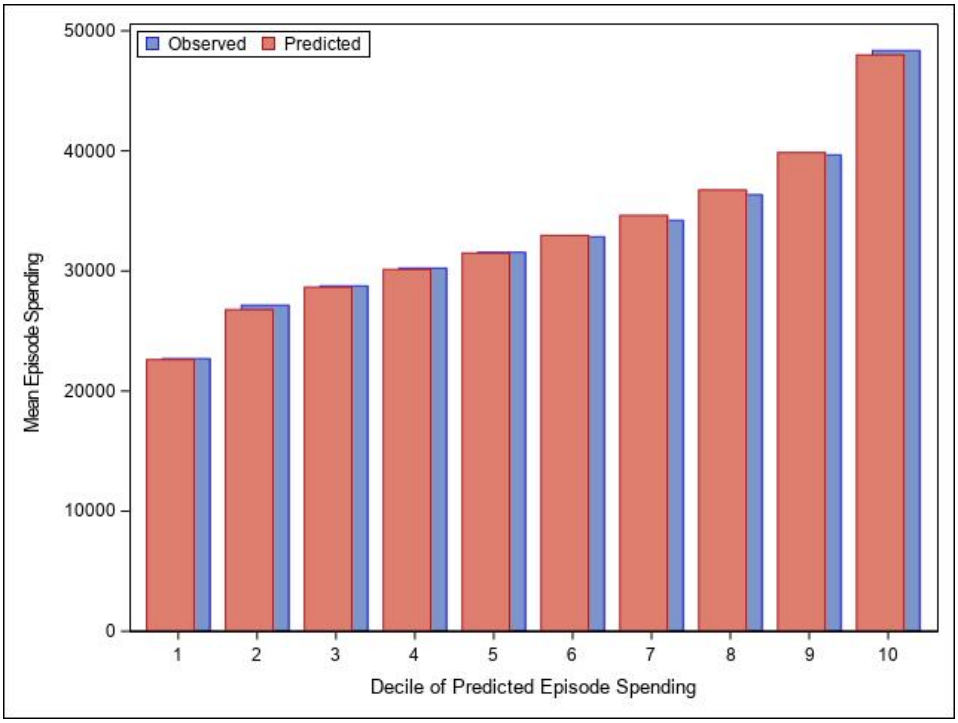
The plots of observed and predicted probabilities for each number of hospital admissions (i.e., 0, 1, 2, …, 10) across quartiles of risk showed that the model performs well across a broad range of risk. In the highest-risk group, we found that the observed and predicted probabilities for zero and one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

differed slightly. However, these differences were small and somewhat expected among the highest-risk group of patients.

Example 2. 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 2. 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 1. 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

Deciles of predicted episode cost	Number of episodes	Observed episode cost	Predicted episode cost	Predicted minus observed cost	Observed / predicted costs
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 providing care in 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.

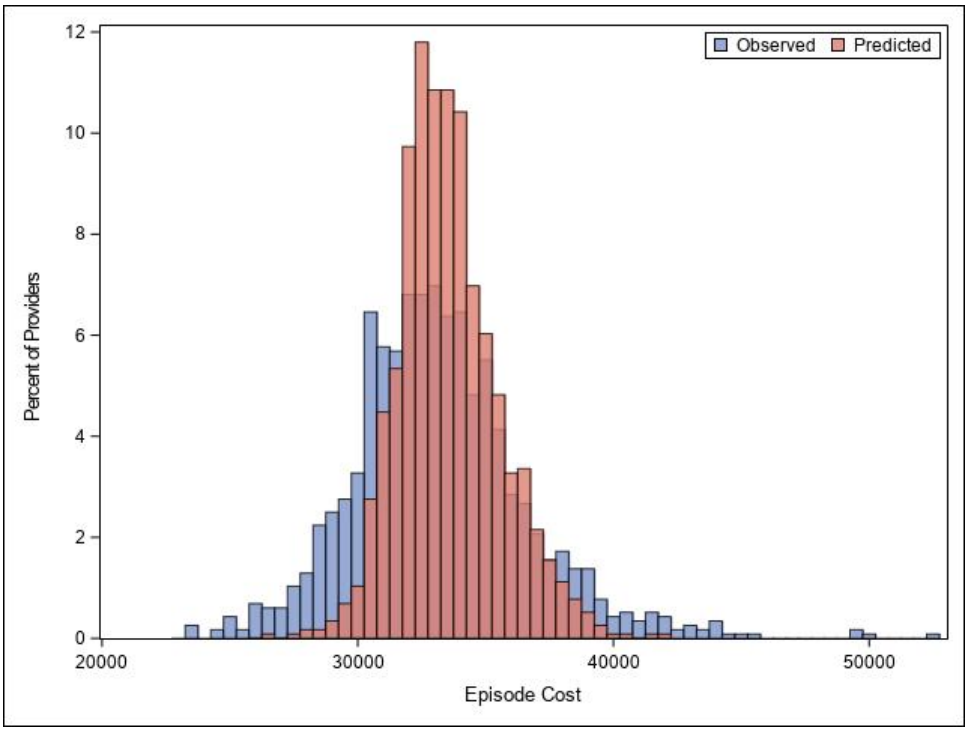
The distribution of facility-level observed and risk-adjusted spending is shown in **Table 12** and **Figure 2.** By considering beneficiary characteristics that are outside of the provider's control, the model compresses the distribution of provider-level spending and decreases its variability. The 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 2. 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 3. Distribution of Provider-Level Observed and Risk-Adjusted Episode Spending



Analysis of Medicare Claims File for IRF FY 2016-2017

Appendix E: Public Comments

Comment

Comment by: Danny van Leeuwen (Health Hats)

The paper buries the issue of health inequities as a risk adjustment problem rather than directly addressing questions of health inequities. I'm not a statistician; I'm a patient caregiver activist. We already know that resources and outcomes vary due to geographic, racial, ethnic, economic, ability, and other variations. The purpose of measurement is not to measure but to motivate and inform improvement, improvement by clinicians, institutions, and communities. Those deciding whether to improve or trying to improve need more than risk adjustment. How can NQF and the measure development industry better inform health equity improvement?

NQF / TEP Response

To directly confront discrimination in American healthcare, NQF is applying an equity lens to every aspect of our work, with the goal of empowering healthcare stakeholders to take meaningful and measurable action to achieve health equity. By striving to consistently apply this lens to our processes, quality measurement and improvement initiatives, and partnerships with local and national stakeholders across the care continuum, NQF will collaboratively and holistically address inequities related to all forms of discrimination.

As part of our five-year strategic plan, NQF will use its unique convening power to collaboratively develop and promote policies and implementation practices advancing the use of data, measurement, and payment models to achieve health equity. This includes addressing quality and measurement gaps in key national health priorities, such as maternal health and access to care. In addition, NQF will build on its previous work addressing data integration to implement a plan that links social determinants of health with clinical data, measurement, and interventions. The integration of data on social determinants and social needs is critical for improving the health of individuals and communities.

NQF will capitalize on partnership opportunities to strengthen the impact of its work. NQF continues to solicit stakeholder feedback on opportunities to address health equity in measurement and implementation, recognizing that health equity is fundamental to all quality improvement efforts. Addressing the wide spectrum of disparities must be considered a key component for successful health outcomes across the nation.

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. On the other hand, the inclusion of social and functional risk factors in risk adjustment models may not make transparent the differences in care outcomes. To mitigate the latter concern, this guidance instructs developers to stratify measure results by key risk factors. Risk stratification is an important tool to deploy in conjunction with risk adjustment to identify healthcare disparities and further promote health equity.

Comment

Comment by: Kidney Care Partners (1 of 5)

Kidney Care Partners (KCP) appreciates the opportunity to comment on NQF's draft report, *Technical Guidance on Developing and Testing Risk Adjustment Models for Social and Functional Status-Related*

Risk Within Healthcare Performance Measurement. KCP is a coalition of 34 organizations comprised of patient advocates, dialysis professionals, healthcare providers, researchers, and manufacturers organized to advance policies that support the provision of high quality care for individuals with chronic kidney disease (CKD) and end-stage renal disease (ESRD). We commend NQF for undertaking this important and timely work.

KCP has long supported efforts to assess and account for social risk factors in the Centers for Medicare & Medicaid Services' (CMS) ESRD Quality Incentive Program (QIP) through adjusters and other mechanisms. As a matter of policy, we have requested that CMS examine measures used in the ESRD QIP and other federal accountability programs to determine how social risk might impact performance and whether adjustment for such factors might improve the measures' ability to differentiate true differences in performance between facilities. As we have noted in our comment letters to the Agency,[1] many measures populating the ESRD QIP and Five Star programs do little to address care disparities and, in some cases, perpetuate inequities. KCP has asked CMS to eliminate or revise such measures to promote health equity and allow the ESRD quality programs to truly empower patients and their care partners.

[1] See, for example, KCP's 2017 letter to CMS on the ESRD QIP.

NQF / TEP Response

No Response Provided

Comment

Comment by: Kidney Care Partners (2 of 5)

As such, KCP appreciates the thoughtful recommendations put forth by the Technical Expert Panel in the draft report. We agree with the TEP that NQF-endorsed performance measures intended for use in accountability applications must provide reliable, valid information about the quality of care provided by the healthcare entity being assessed. Measures used to determine financial penalties, in particular, "should be affected as little as possible by factors other than quality of care, such as patient characteristics already present at the start of care." To avoid unfairly penalizing providers for patient and community characteristics beyond their control, risk adjustment and/or stratification can be critical to the design of effective value-based payment programs. While adjustment for social risks has remained controversial for fear of masking disparities or tacitly forgiving lower quality of care for socially marginalized patients, an increasing evidence base suggests that a failure to appropriately consider such variables may in fact exacerbate existing and ingrained sociodemographic, economic, and geographic disparities by disproportionately penalizing the safety-net facilities caring for our most vulnerable patients.[1],[2]

[1] Joynt KE, Jha AK. Characteristics of hospitals receiving penalties under the Hospital Readmissions Reduction Program. JAMA. 2013;309 (4):342-343.

[2] Medicare Payment Advisory Committee. Chapter 4: Refining the hospital readmissions reduction program. In: Medicare Payment Advisory Committee. Report to the Congress: Medicare and the Health Care Delivery System. Washington, DC: MPAC; 2013:91-116.

NQF / TEP Response

No Response Provided

Comment

Comment by: Kidney Care Partners (3 of 5)

Nevertheless, we have a number of concerns with the TEP's guidance in the report. First, we highlight that the recommendations are at odds with the Office of the Assistant Secretary for Planning and Evaluation's (ASPE) advisory reports to Congress on the use of social risk factors in Medicare's Value-Based Purchasing Programs,[1] most notably on the issue of whether outcome measures should be adjusted for social risks. The TEP dismissed this incongruity because NQF has historically taken an "agnostic" stance to measure use and is here simply providing a framework to meet the standards for NQF endorsement. We believe this is a distinction without a difference, as it cannot be denied that measure use in accountability programs has now become inextricably intertwined with NQF endorsement. We further posit that this conspicuous discrepancy is fundamentally at odds with the NQF mission of reconciling redundant and incompatible guidelines, standards, and measures offered by various healthcare quality improvement organizations and agencies. Despite the underlying premise of the report—to facilitate consistency in the evaluation of risk adjustment models within performance measures—we are concerned that such striking inconsistencies paradoxically risk perpetuating ambiguity and confusion among measure developers and other stakeholders.

[1] See, for example, HHS Office of the Assistant Secretary for Planning and Evaluation (ASPE). Report to Congress: Social Risk Factors and Performance in Medicare's Value-Based Purchasing Programs. March 2020.

NQF / TEP Response

Comment (3 of 5)

NQF takes into consideration the guidance in the ASPE report. Like the ASPE report, this Technical Guidance acknowledges the importance to risk-adjust certain outcome and cost/resource use measures, where appropriate. This Technical Guidance also agrees that there is a need to improve risk adjustment overall to meet the demands of a changing healthcare landscape. To that regard, this Technical Guidance describes a framework of minimum standards that developers should consider for social and/or functional risk adjustment within quality measurement.

However, as an independent standards-setting organization, NQF convenes expert stakeholder groups to make independent recommendations and assessment. One important distinction between the ASPE report and the recommendations in the NQF's Technical Guidance report lies in the intent behind both the reports. The ASPE report's recommendations are primarily intended for Medicare's Value-Based Purchasing (VBP) program, while NQF makes standard-setting recommendations on measures for use in quality improvement and accountability applications that include Medicare's VBP program and other private sector programs. To the regard, the minimum standards outlined are to provide measure developers with the necessary tools needed achieve NQF endorsement, respective to social and/or functional risk adjustment. Although NQF may not control how measures are implemented or used (e.g., within VBP applications), it is important to signal that program polices have an impact on accountable entities caring for populations with social and/or functional risk. Outcome and cost/resource use

measures need to consider the conceptual relationship of social risk factors at the start of care, provider actions, the provider locus of control to have an impact on the social risk factor, and the intended use (or the incentives/resources allocated to have an impact on the social/functional risk factors) in determining whether or not social and functional risk adjustment is appropriate.

Comment

Comment by: Kidney Care Partners (4 of 5)

It is also unclear to us what the TEP is recommending vis-à-vis risk stratification. The seven Minimum Standards put forth in the report seem to indicate that social risk adjustment is now an endorsement requirement for outcome measures and that risk stratification should be conducted in conjunction with risk adjustment to ensure the adjusted measure is able to identify healthcare disparities. Elsewhere, and ostensibly at odds with the Minimum Standards, it is noted that stratification can be an appropriate alternative to risk adjustment, subject to the developer's assessment of the role of social and functional risk factors in the context of the specific intended use of the measure. We request additional clarification on this point, and we urge NQF to allow developers discretion in this regard. Stratification may indeed be the most appropriate approach to social risk in some outcome measures, allowing providers and other healthcare stakeholders to identify and prioritize differences in care, outcomes, and experiences across different sociodemographic groups and to develop and implement equity-focused practices to better address disparities and understand the experiences of patients from marginalized communities.[1] Such insights would be obscured if the same measures were instead adjusted for social risks.

[1]See Advancing Health Equity. "Using Data to Reduce Disparities and Improve Quality." https://www.solvingdisparities.org/sites/default/files/Using%20Data%20Strategy%20Overview%20Oct.%202020.pdf (accessed June 22, 2021).

NQF / TEP Response

Comment (4 of 5)

The minimum standards outlined are to provide measure developers with the necessary tools needed for NQF endorsement, respective to social and/or functional risk adjustment. Although NQF may not control how measures are implemented or used (e.g., within VBP mechanisms), it is important to signal that program polices have an impact on accountable entities caring for populations with social and/or functional risk. In terms of risk stratification, if a performance measure includes social risk variables in its risk adjustment model, the measure developer must provide the information required to stratify a clinically-adjusted-only version of the measure results for those social risk variables. This information should include the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically adjusted version of the measure when appropriate. Furthermore, stratification is not a risk adjustment approach but simply a way of presenting subgroup-specific data, typically unadjusted, in order to demonstrate differences in outcomes among various groups. This is not an alternative to the risk adjustment approaches discussed above but rather a parallel approach, which should always be presented along with risk-adjusted results when considering certain variables, such as SES/race—by using both approaches, you maximize information and minimize the risk of missing or obscuring important differences. NQF seeks further

input from the TEP regarding risk stratification requirements. Based on the TEP's feedback, NQF's team will update the Technical Guidance report to provide further clarification.

NQF would also like to emphasize that the NQF-convened TEP for this project is providing scientific guidance that will later need to be considered for updates into the NQF endorsement process. NQF seeks further input from the TEP regarding risk stratification requirements. Based on the TEP's feedback, NQF's team will update the Technical Guidance report to provide further clarification.

Comment

Comment by: Kidney Care Partners (5 of 5)

Finally, NQF indicates that in formulating the recommendations laid out in the report, it considered the potential burden for measure developers and that the increased requirements might create barriers to measure development. As indicated in the Core Principles supporting the report, the identified statistical approaches "are not intended to be overly prescriptive, as to limit the use of novel methods or to add significant burden to measure developers." However, the report then prescribes in specific detail what types of testing methodologies are—and are not—acceptable. For example: "Simple bivariate and multivariable tests alone should not determine whether a social or functional risk factor is included in the risk model . . . Additional calibration and discrimination tests of the risk adjustment model in subpopulations specific to the measure should also be done . . . Developers should use caution in that changes in model discrimination, such as c-statistics, may not be enough to inform a decision to include an additional social and/or functional risk factor in the model specification." We fear that such rigid recommendations, which will likely be adopted wholesale by NQF's Scientific Methods Panel and Standing Committees as criteria against which to evaluate measures for endorsement, will indeed stifle innovation—particularly among small developers with limited resources. As indicated in NQF's prior work in this area, developing social risk strategies for performance measures is an iterative process involving empirical analyses and multiple decisions to arrive at a final procedure. "There is more than one appropriate way to accomplish adjustment, . . . [and] NQF should not be prescriptive regarding methods for adjustment or specific SDS variables."[1] We urge NQF to heed its own advice in this regard and to grant measure developers the flexibility to use the best methods indicated in a particular situation.

KCP again thanks you for the opportunity to comment on this important work.

[1] National Quality Forum (NQF). Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors Technical Report. August 15, 2014.

NQF / TEP Response

Comment (5 of 5)

Measure developers, stewards, and program implementers have long expressed a need for technical guidance and standardization in developing, testing, and evaluating risk adjustment models that account for social and/or functional risk. Therefore, NQF would like to clarify that the intent of this Technical Guidance report is to provide several illustrative examples of empirical testing approaches that developers may consider (Appendix D in the Technical Guidance report). The TEP also recommended that although the empirical testing is not deterministic, developers should examine that evidence in conjunction with the conceptual model. Developers should also describe the statistical methods used and the results and interpretation of the analyses. Developers should be transparent about their approach and their interpretation of the results.

Lastly, NQF continues to acknowledge that current measure evaluation criteria remain "agnostic" to measure use. The current NQF endorsement criteria of use and usability are intended to ensure that endorsed measures can be used in quality improvement and/or accountability applications. The intent of this guidance is to inform new approaches without constraining the TEP of the current NQF criteria. NQF recognizes that more work is needed to operationalize further, and NQF hopes to accomplish that in an Option Year, if awarded.

Comment

Comment by: American Association on Health and Disability

The American Association on Health and Disability (AAHD) (www.aahd.us) is a national, nonprofit organization of public health professionals, both practitioners and academics, with a primary concern for persons with disabilities. The AAHD mission is to advance health promotion and wellness initiatives for persons with disabilities. AAHD is specifically dedicated to integrating public health and disability into the overall public health agenda.

The Lakeshore Foundation (www.lakeshore.org) mission is to enable people with physical disability and chronic health conditions to lead healthy, active, and independent lifestyles through physical activity, sport, recreation and research. Lakeshore is a U.S. Olympic and Paralympic Training Site; the UAB/Lakeshore Research Collaborative is a world-class research program in physical activity, health promotion, and disability linking Lakeshore's programs with the University of Alabama, Birmingham's research expertise.

Our comments reinforce and support three of the draft report observations and recommendations.

> RE: 9 recommended core principles – pages 7 & 8. We reinforce the overriding importance of core principle #2 – disparities should be identified and reduced
>
> RE: Medicare and Medicaid dual eligibility as an important proxy for the underserved and challenged populations – page 12. As a 2012-2017 member of the NQF Workgroup on persons dually eligible for Medicare and Medicaid, there is an abundance of data and information on the burden faced by these persons and the systems and providers that serve them. We reinforce the importance of this recommendation.

Recommended Minimum Standard – page 12: "At a minimum, developers should consider age, gender, race/ethnicity, urbanicity/rurality, Medicare and Medicaid dual eligibility, indices of social vulnerability, and markers of functional risk (frailty, ADLs, IADLs)." We reinforce this minimum standard. We also bring to NQF's attention: 2000-2021 recommendations of the Consortium for Citizens with Disabilities (CCD), Disability and Rehabilitation Research Coalition (DRRC); and Trust for America's Health (TFAH) led public health reporting collaborative: demographic data collection, analysis, and public sharing should "include in every measure by emphasizing the importance of stratification and cross-tabulation of data by race, ethnicity, disability status, age, sex, sexual orientation, gender identity, race, ethnicity, primary language, rural/urban environment, and service setting for all core measures."

NQF / TEP Response

No Response Provided

Comment

Comment by: Federation of American Hospitals

The Federation of American Hospitals (FAH) appreciates the focus of this draft report to further guide measure developers on the minimum standards for the consideration of social and/or functional status-related risk factors in a measure's risk adjustment approach.

The FAH strongly supports the expectation that measure developers should consider a wide set of data sources and variables as outlined in the second standard. As we anticipate that new and novel data will become available over time, we encourage NQF to ensure that this list of sources and variables is updated frequently.

The FAH recommends that NQF consider including a minimum standard that requires measure developers to consistently provide data on how the inclusion of one or more of these factors may or may not shift the performance of the accountable entities. When these data have been provided in previous submissions, it has enabled us to make determination on the degree to which scores could be positively or negatively impacted, and we believe that it directly relates to whether potential unintended consequences may result based on the measure's use.

The FAH also encourages NQF to consider whether a minimum standard should be added that encourages developers to consider other strategies beyond the usual approach of "adding on" social and/or functional status-related factors after clinical variables such as multilevel models or testing of social factors prior to clinical factors as suggested in the Evaluation of the NQF Trial Period for Risk Adjustment for Social Risk Factors report.

Thank you for the opportunity to comment.

Reference:

National Quality Forum. Evaluation of the NQF Trial period for Risk Adjustment for Social Risk Factors. Final report. July 18, 2017. Available at:

http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=85635. Last accessed December 18, 2018.

NQF / TEP Response

Thank you for taking the time to review the Technical Guidance report and providing your comment. NQF agrees that although standardized empirical testing approaches are beneficial, the intent of this guidance is not to be prescriptive to the types of empirical testing that the developer should conduct. Empirical analysis assessing the impact of a specific social or functional risk factor on the distribution of measured entities' performance may not be deterministic in the decision of whether to include the factor in the risk model, but rather relies on the conceptual model outlined by the TEP. This guidance purposes to provide several illustrative examples of empirical testing approaches that developers may consider (Appendix D in the Technical Guidance report). The TEP also recommended that although not deterministic, developers should examine the empirical evidence in conjunction with the conceptual model. It further asks the developers to describe the statistical methods used and the results and interpretation of the analyses, which leads to the decision of whether or not to select social and/or functional risk factors for risk adjustment. Lastly, the guidance notes that the developers should be transparent about their approach and their interpretation of the results.

Comment

Comment by: National Committee for Quality Assurance (NCQA)

With regards to the comments about what is meant by the calibration section—generally, I (Rachel) interpret the intent as "is it predicting as well within the subgroup"—part of this goes to causal model, but I think part of it begs the question—do certain subgroups deserve models that predict specifically to their population? I think much of the confusion in the insulin example comes from expanding this calibration question outside the social/functional risk use case, since and what is modifiable in the causal pathway. This guidance is for social and functional risk—is it modifiable? Is it in causal pathway as a level for change? In some cases, yes, and in others, no. One could argue SES (income/education/etc.) is non-modifiable, and not in pathway of intervention, so the model should be designed to predict for effect modification vs. "showing gap." This is the "art" of the discussion though. A bit more clarity on this, especially since it's a relatively new way of approaching, would probably be helpful.

NQF / TEP Response

No Response Provided

Comment

Comment by: NCQA

Great work! Some comments:

> The examples are helpful. The exposition on EDAC was very good about explaining pathways.
>
> How the measure is intended to be used is indeed important, as there might be different unintended or intended consequences for something that is used for public reporting vs. VBP. However, there is often mission creep whereby a measure published for one purpose gets repurposed for something else. HCCs are a good example of this, as is the Charlson score. Suggest more explicit acknowledgements of this in the document.

The report clearly has providers, provider groups, and hospitals most in mind rather than plans. That makes it a bit tricky to apply to our work at NCQA because plans might have different levers available to them. Different variables might be considered modifiable for a plan vs. a provider. I'm not sure a whole other guidance document, this time with plans in mind, would be advisable, but I'm wondering about how to translate. Suggest perhaps a specific review to: 1) make sure entity specific language isn't used when recommendation is general and 2) make sure the point around how causal path/modifiability varies by accountability model is articulated.

On the standard that calibration should be conducted with subpopulations, the scope of this remains unclear. I was wondering:

> By "conducted", do you mean that we should assess and report calibration in subpopulations? Or that we should actually make sure the model is well calibrated across different subpopulations?
>
> Some of the material in this guidance document suggests the latter (e.g. guidance that measure developer should show that a model does not mispredict the outcome systematically for certain subgroups.
>
> But what if that misprediction is in some sense the point of having the measure? For example, in [one draft measure for a diabetes outcomes] the model is very uncalibrated for each subgroup of (insulin users, insulin nonusers). However, the advisors have argued that managing insulin is one major mechanism for intervention, so a miscalibration by insulin status simply reflects the variable that the measure is intending to modify.
>
> A different example, if a model mispredicts outcomes for Black versus White patients, does that mean the measure should adjust for race? Or maybe you want the measure not to adjust for race so that entities can be held accountable for the worse health outcomes experienced by Black patients on average.

NQF / TEP Response

Thank you for taking the time to review the report and providing your comment. In regard to your comment surrounding measure use, NQF evaluates measures for quality improvement and accountability applications. NQF recognizes that the conceptual model should outline the locus of control of the accountable entity and specific intended use of the measure when determining the appropriateness of social and functional risk adjustment. NQF will seek further input from the TEP on the importance of the intended use for the measure in the conceptual model. Based on the TEP's feedback, NQF proposes addition of a statement in the intended use section to ask the developers to include the applications where the measure should not be used. NQF will seek to consider operationalizing this guidance in the endorsement criteria within the Option Year, if awarded.

Additionally, to provide further clarification on the accountable entities, the guidance notes that accountable entities can include clinicians, health plans, and health systems/hospitals.

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. The TEP agrees that there may be statistical reasons to consider a separate model for certain population subgroups. For example, if you have data suggesting that one SES/racial group has much different outcomes than another or that the association of other covariates with outcomes is much different for that subgroup, then you could argue that they would be better served by their own model. Calibration and other performance metrics would probably be better but at the cost of losing generalizability of the model to other populations. Another reason for a separate model would depend on the policy goal of the measure developer and measure implementer. Again, NQF recognizes that more work is needed to operationalize this further, and NQF hopes to accomplish that in an Option Year, if awarded.

Comment

Comment by: Yale CORE

#3597 is now NQF-endorsed, as of the June 2021 CSAC meeting.

RTI is the developer for the ACO version of #1789, so it would be good to include them in the document and let them know that their content is included in the report (if you have not already).

The table on page 49 is not rendering correctly in the pdf.

NQF / TEP Response

Thank you for the comments. The Technical Guidance has been updated to reflect these changes.