NATIONAL QUALITY FORUM

Driving measurable health improvements together

Best Practices for Developing and Testing Risk Adjustment Models: Option Period 1 Public Comments

https://www.qualityforum.org

1099 14th Street NW, Suite 500 | Washington, DC 20005 | M 202.783.1300 F 202.783.3434

Table of Contents

Purpose

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

NQF staff have included all comments that were received in this memo; it contains the commenter's name, comment, themes derived from the comment, sections of the Technical Guidance impacted, a proposed response and/or proposed adjudication, and questions for the Technical Expert Panel (TEP). A proposed response answers questions from the respective comment for the TEP's consideration. However, the proposed responses do not suggest edits to the Technical Guidance. Rather, a proposed adjudication, if provided, contains proposed edits to the Technical Guidance for the TEP to consider. For wayfinding in the document, NQF staff separated comments to contain paragraph numbers. Readers can match the proposed response and/or adjudication with the corresponding paragraph number in the comments.

Comments received are shared with the TEP prior to the October 24, 2022, public meeting. During the meeting, the proposed responses and adjudications will be presented for the TEP's consideration and discussion. The TEP will be given an opportunity to suggest changes to the responses, additional adjudications, or agree with the suggested responses and/or adjudications.

Prompt 1: Please provide feedback on the recommendations for stratification (pg. 40-42), including the appropriateness of the stratification methods, the reporting methods (e.g., establishing cut points for strata), and the stratification variables.

Comment #1

Commenter: Doris Peter, Yale University School of Medicine

Date Comment was Submitted: 9/21/2022

TEP Discussion Time Allotted: 10 minutes

Comment

1. Thank you for the opportunity to comment. The guidance is unclear – in some places in the report, it states that stratification should submitted only for measures that are adjusted for social/functional risk. In other places it does not say that. Can you please clarify?
2. We are concerned that these stratification requirements could be a significant burden to developers. Developing the proper stratification approach can take a substantial amount of time and resources, and to get it right would require meaningful and iterative engagement with stakeholders. This could be a multi-year project for some measures and some developers may not have sufficient funding available.

Theme

Providing clarity
Burden to developers
Stratification guidance

Sections Impacted

Considerations for Determining the Final Risk Adjustment Model
Stratification
Minimum Standards

Proposed Response

1. Thank you for the comment. The intent of the stratification recommendation within this Technical Guidance is for stratification to be considered independently of decisions regarding risk adjustment. Specifically, *"as mentioned above, the purpose of risk adjustment in measurement is to create measures that fairly compare quality of care between measured entities. Stratification enhances transparency to highlight known areas where disparities exist, which would be researched as part of the construction of the conceptual model"* (pg. 41). Stratification should be used to enhance transparency and known areas where disparities exist. These disparities would have been researched as part of the construction of the conceptual model.
2. NQF recognizes that stratification may add requirements to the NQF endorsement process, but these requirements will significantly advance uncovering disparities in care and further promote health equity. The technical guidance does acknowledge limitations, for example, *"for this [stratification] subgroup analysis and for the groups identified in the conceptual model, a measure developer's capabilities may also depend on available data and sample sizes within subgroups. If this is the case, measure developers should demonstrate why the analysis was not feasible"* (pg. 41).

Proposed Adjudication

None

Questions for the TEP

- Does the TEP agree with the proposed responses?
- Does the Technical Guidance need additional changes to improve clarity?

Comment #2

Commenter: Joyce Lam, Acumen, LLC

Date Comment was Submitted: 9/21/2022

TEP Discussion Time Allotted: 10 minutes

Comment

1. Minimum Standard #7 calls for stratification of measure scores by social risk factors to highlight within-provider and across disparities. An important point of clarification is whether the stratified reporting is intended to be done prior to or after risk adjusting for SRFs, if the measure does adjust for SRFs? If the latter, there will likely be little or no residual difference.
2. In addition, the purpose of these analyses in the evaluation process is not clear. Questions that the technical guidance could better answer are ones such as:
    a. If measure scores differ across high SRF and low SRF subpopulations, what is the implication of that for the evaluation process?
    b. Will the evaluation panel then require the measure to be reported in a stratified manner? If so, how will NQF address the fact that many measure developers are separate from measure reporting/production contractors and have no role in measure reporting?
    c. If some providers have zero or limited "within" disparities while others do not, what is the implication for including the SRF in risk adjustment? Does this mean the measure is more or less "valid"?
3. We recognize that examining stratification of measure scores provides useful and interesting information for policy makers. But without answering concrete questions such as the above, Minimum Standard #7 poses the danger of adding significant burden and cost to measure developers and CMS, with no clear goal or standard in relation to the NQF evaluation process.

Theme

Providing clarity
Stratification guidance
Requirements for Consensus Development Process

Sections Impacted

Technical Guidance Overview
Considerations for Determining the Final Risk Adjustment Model
Stratification
Minimum Standards
Conceptualizing the Model

Proposed Response

1. Thank you for your comment. Stratified reporting is intended to be done prior to risk adjustment using either 'unadjusted' or 'clinically risk adjusted' data (i.e., not yet adjusted for social or functional risk). As noted in the guidance, *"depending upon the specific application,*

results stratified by social and functional risk subgroups may use unadjusted or clinically risk-adjusted (e.g., age, comorbidities) data, but not the same stratification variables that have already been used for social or functional risk adjustment." (pgs. 40-41).

2. Thank you for your comments and questions regarding the evaluation of stratification approaches.
    a. If measure scores differ across high SRF and low SRF subpopulations, developers should acknowledge this difference but in and of itself this does not determine whether an SRF should be included or not. First, the developer should understand through the conceptual model how the particular SRF influences the measured outcome and the decision to include or not include the SRF should be informed primarily by the conceptual model. As in the Technical Guidance, "*the developer reviews the conceptual model and empirical results and considers the full range of implications in deciding whether to adjust for social and/or functional risk factors. 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"* (pg. 17).
    b. NQF recognizes that neither developers nor NQF are able to specify requirements for reporting measures. However, as stated in the Technical Guidance, *"Broadly, there are several areas of the NQF measure information specifications and the validity component of the NQF Scientific Acceptability of Measure Priorities endorsement criterion that will require updating. These areas include…specifications on how the measures should be stratified if a social or functional risk factor is included in the final risk adjustment model"* (pg. 44).
    c. Thank you for your comments regarding how variation in SRFs contributes to the validity of the measure. The importance of variation may depend on several factors such as the conceptual model, or other evidence gathered from the literature review (e.g., showing that the SRF can be mitigated). As stated in the Technical Guidance, "*Measure developers must examine whether and how much it is within the accountable entity's locus of control to mitigate the risk factor. Developers can demonstrate an accountable entity's ability to meaningfully influence a factor by citing the primary literature, public reports, and case studies, and/or by conducting empirical analyses, as described in the Locus of Control section below, to determine the variation and degree of impact of a social risk factor to a measured outcome…Variation suggests that some entities can successfully mitigate a factor, as reflected by observed differences in outcomes, and that at least for a subset of entities the factor is within their locus of control."* (pg. 25). The TEP should discuss the implications of adjusting or not adjusting when "within" entity variation is zero for some and non-zero for others.

Proposed Adjudication

None

Questions for the TEP

- What is the implication of adjusting or not adjusting when "within" entity variation is zero for some and non-zero for others?
- Does the TEP agree with the proposed responses?
- Does the Technical Guidance need additional changes to improve clarity?

Prompt 2: The minimum standard best practices developers should follow have been updated based on stakeholder feedback. Please provide feedback on whether the minimum standards (pg. 19-20) need additional specifications or clarifications.

Comment #3

Commenter: Joyce Lam, Acumen, LLC

Date Comment was Submitted: 9/21/2022

TEP Discussion Time Allotted: 15 minutes

Comment

1. **Minimum Standard #1** In some cases, Minimum Standard #1 appears to be in conflict with **Minimum Standard #2**, such as situations where the literature has identified a different set of relevant SRFs or functional risk factors than in Minimum Standard #2. For example, the Andersen-Newman framework of health services utilization, which does not explicitly call for urbanicity/rurality but uses other indicators of social structures, may satisfy Minimum Standard #1 but not Minimum Standard #2. In other words, adhering to Minimum Standard #2 may force developers to discount the findings from the literature that point to a particular set of SRFs that are conceptually linked to the outcome.
2. For **Minimum Standard #1**, there are hundreds of possible SRFs that could theoretically have a relationship with the outcome of interest. An exhaustive accounting of every single potentially relevant SRF, and the associated literature, is substantially burdensome. This is unlikely to be what NQF intends; if so clearer criteria as to what factors the conceptual model should consider would be critical to state.
3. **Minimum Standard #3** This appears to set the presumption of measure invalidity. That is, it shifts the burden of proof to the developer to show that a measure is valid when there is no data for potential SRF variables. We suggest this be clarified.
4. **Minimum Standard #5** This minimum standard represents a significant burden to developers, as well as evaluation panels. For example, if measures use the HCC model to risk adjust for clinical characteristics, there will be approximately 100 variables to analyze individually. Page 36 states that these results will not be used to determine whether a risk factor should be included in the model. Therefore, this minimum standard is unnecessary when considering its substantial burden for developers and evaluation panels. If NQF's intent is instead to mandate descriptive analysis only for SRFs and functional status items in the conceptual model, then we would suggest clarifying that in the minimum standard wording. Assuming this is the intent, providing clearer guidance on what SRFs should be included in Minimum Standard #1 will help contain the burden associated with Minimum Standard #5.

Theme

Providing clarity
Burden to developers

Sections Impacted

Minimum Standards
Conceptualizing the Model
Identifying and Selecting Potential Data Sources and Variables
Empirically Testing Risk Factors

Proposed Response

1. Minimum standard #1 requires no edits. Conceptual models should be formed based on the literature. Factors in Minimum Standard #2 should supplement what is identified through the literature search conducted in Minimum Standard #1. These minimum standards focus on selection of factors for consideration in the conceptual model and are not required for inclusion in either the final conceptual model or for the testing or final implementation of the risk adjustment model.
2. Regarding the number of SRFs required to be considered, the Technical Guidance notes that "*For the first step, measure developers should explore the broad list of factors that might have an impact on the outcome. This Technical Guidance is not overly prescriptive about requirements for evidence so as not to burden measure developers or dampen development in unmeasured areas of disease and healthcare. At a minimum, evidence for inclusion in the conceptual model should include a literature review to identify the most important and plausible factors. A systematic review is not required, and it may be advisable to turn to the literature for evidence as well. Developers should also attempt to supplement the literature review with expert opinion to identify and uncover gaps in the literature…There are a number of social and functional risk variables that must always be considered in the conceptual model for outcome and cost/resource measures. This best practice was agreed upon by the TEP, which identified a minimum set of factors that are commonly used and analyzed by developers….: age, gender, race, ethnicity, an indicator of urbanicity/rurality, indicator of poverty (such as Medicare and Medicaid dual eligibility and income), indices of social vulnerability (such as the Area Deprivation Index [ADI] or the Agency for Healthcare Research and Quality [AHRQ] SES Index score), and indicators of frailty and disability (such as ADLs, vision, hearing, cognitive impairment, eligibility for disability programs, etc.) (Minimum Standard #2). Therefore, all relevant demographic factors, clinical risk factors, social and functional risk factors, and patient preferences related to the outcome of interest should be considered for inclusion in the conceptual model. The conceptual model depicts all relevant factors, regardless of whether they will be used in the final risk adjustment model or whether data can be operationalized in the full measured population"* (pg. 26-27).
3. Regarding Minimum Standard #3 and the concerns around the presumption of measure invalidity. This is correct. It is the developer's responsibility to analyze the impact of a variable that was identified as important in the conceptual model. The technical guidance states that, *"at a minimum, developers should examine previously published evidence and should attempt to estimate the directionality of the bias for the factor of interest by using other studies. This may be achieved by reviewing the literature to determine how a risk factor might affect subsets of the accountable entities…If there is a high degree of unevenness across accountable entities, the developer may choose to exclude those accountable entities that have a large proportion of the risk factor from the overall group being measured, or the developer may consider including proxy variables of social and/or functional risk in the risk adjustment model based on prior research. For the latter, the relevance of these proxy variables should be empirically appropriate for the measured outcome of interest (see the section titled Operationalizing Variables for Risk Adjustment for more information about the use of proxy variables)"* (pg. 31-32).

Proposed Adjudication

2. Minimum standard #2 requires a small edit to clarify that consideration of these factors is for their potential inclusion in the conceptual model before it is finalized. "*Developers must, at a minimum, consider age, gender, race, ethnicity, an indicator of urbanicity/rurality, indicator of poverty (such as Medicare and Medicaid dual eligibility and income), indices of social*

vulnerability (such as the Area Deprivation Index [ADI] or the Agency for Healthcare Research and Quality [AHRQ] SES Index score), and indicators of frailty and disability (such as ADLs, vision, hearing, cognitive impairment, eligibility for disability programs, etc.) ***for inclusion during the process of developing*** *the conceptual model"* (pg. 19, 28).

4. Response for clarification of wording on minimum standard #5 – this minimum standard is specific to descriptive statistical tests, as the intent is to mandate descriptive analysis for SRFs and functional status in the conceptual model to understand the populations served and potential disparities. "*This uneven distribution may result in a bias for a subset of accountable entities that disproportionately serve patients with the social risk factor. If there is a high degree of* ***difference in SRF distribution*** *across accountable entities, the developer may choose to exclude those accountable entities that have a large proportion of the risk factor from the overall group being measured, or the developer may consider including proxy variables of social and/or functional risk in the risk adjustment model based on prior research* (content has been edited out of the original). *For the latter, the relevance of these proxy variables should be empirically appropriate for the measured outcome of interest (see the section titled Operationalizing Variables for Risk Adjustment for more information about the use of proxy variables)"* (pg. 31-32).

Questions for the TEP

- Does the TEP agree that the descriptive statistic requirement is necessary, given the balance between developer burden and identifying variation?
- Does the Technical Guidance need additional changes to improve clarity?
- Does the TEP agree with the proposed responses?
- Does the TEP agree with the proposed adjudication to the Technical Guidance?

Prompt 3: The sets of factors and variables in minimum standards #2, #6, and #7 have been updated to reflect stakeholder and TEP input. Please provide feedback as to whether there are challenges to operationalizing these minimum standards (pg. 19-20).

Comment #4

Commenter: Joyce Lam, Acumen, LLC

Date Comment was Submitted: 9/21/2022

TEP Discussion Time Allotted: 15 minutes

Comment

1. **Minimum Standard #2** This minimum standard calls for at least 8 separate variables that developers must "consider". It is not immediately clear how developers can demonstrate that they have fulfilled this minimum standard, such as through additional testing that will be required on the submission forms, simply through the developer's internal discussion, or somewhere in between. Without guidance for evaluation panels, there will be ambiguity for developers and panels alike.
    a. We suggest clarifying whether the intent is to use gender or biological sex (or race or ethnicity), how they are captured from data sources, their relationship with a health outcome, and the implication if they are used in risk adjustment.
2. **Minimum Standard #6** The guidance on empirical testing calls for predictive ability and discrimination with and without the SRFs and functional risk factors from the list in Minimum Standard #2, and calibration results across race/ethnicity, urbanicity/rurality, an indicator of poverty, and an indicator of disability. At the same time, the guidance appears to imply that none of the three criteria (predictive ability, discrimination, and calibration) are determinative in considering whether to risk adjust for SRFs and functional risk. More generally, there is no guidance for evaluation panels about how they should use these empirical results.
    a. Given this, it is not clear if there is any additional value for the purposes of NQF evaluation in conducting these tests, beyond the usual calibration/discrimination tests. The lack of a clear tie between empirical testing results and evaluation standards suggests that, while this minimum standard may lead to interesting and important analyses for policy makers and academics, it is not necessary for the NQF evaluation process. Limited predictive ability or calibration results showing mismatches between observed and expected outcomes are not always (and may not even often be) signs of an invalid risk adjustment model. For example, calibration results for African-American Medicare enrollees may indicate under-prediction of adverse outcomes for these beneficiaries. What to do with this finding is not clear. It does not necessarily mean the risk adjustment model underlying the measure is invalid. In fact, the risk adjustment model could be valid and the calibration exercise is revealing that there is systematically poor care for African-American enrollees that must be addressed through improved provider performance.
3. **Minimum Standard #7** Please see our comments on stratification.

Theme

Provide Clarity

Sections Impacted

Minimum Standards

Proposed Response

1. Thank you for your comments regarding how to demonstrate whether developers "considered" the factors adequately. The literature review and narrative description of the conceptual model and the process for developing the conceptual model are inputs that the Technical Guidance identifies for the developer's consideration when examining factors for potential inclusion in the conceptual model. These factors should be considered in the hypothesized relationships within the conceptual model. The TEP's intention was not to be overly prescriptive about whether gender or biological sex is appropriate as it may depend on the outcome being measured and the data available. The TEP acknowledges that depending on how risk factors are captured from data sources, their relationship with a health outcome, and the implication if they are used in risk adjustment are all influencing factors in the decision to include or exclude a factor, such as gender or biological sex. The developer must demonstrate their consideration of these factors and their effects.
2. Regarding comments on Minimum Standard #6, a developer may consider these tests by reporting on any associations in the literature, supplemented by expert opinion, and the magnitude of the results. As stated in the Technical Guidance, the purpose of the empirical tests is to test the relationships hypothesized in the conceptual model "…*there are several empirical testing methods that may be used to test the relationships described within the conceptual model*" (pg. 17). The Technical Guidance is not prescriptive about which results must be used. Additionally, the TEP should review and discuss how to use these empirical results to add guidance. NQF can then use the TEPs discussion as the basis of future changes to the measure evaluation criteria.

Questions for the TEP

- The TEP should review and discuss how to use these empirical results to add guidance.
- Does the TEP agree with the proposed responses?
- Does the Technical Guidance need additional changes to improve clarity?

Comment #5

Commenter: Danielle Lloyd, America's Health Insurance Plans (AHIP)

Date Comment was Submitted: 9/21/2022

TEP Discussion Time Allotted: 10 minutes

Comment

1. AHIP appreciates the consideration of mediators of social risk factors, and we recognize the balance to determine what aspects are part of providing high-quality care and what is outside of the control of a measured entity. However, we are concerned that these mediators may not be equally distributed as the scope of a particular problem can differ by location and communities have varying capabilities and capacities to respond. Moreover, states have varying policies and funding available that influence how communities can respond to social needs.
    a. For example, food insecurity could theoretically be mediated by a referral to the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) program, but the actual impact of that benefit could vary based on how a state has implemented the program as well as if the person lives in a food desert and is able to access healthy foods. As another example, some states have more robust home and community-based services programs that can provide more information and access to consumers by providers and plans to better understand social factors and the best approach for mitigation. This variation may limit or enhance a measured entity's ability to mediate the impact of potential social risks on health outcomes.
2. While we recognize that some entities have demonstrated an ability to successfully mitigate some social risk factors, NQF-endorsed measures are intended to be national standards and developers should consider the impact of regional variation in resources. We suggest adding text on page 25 to address this concern.
    a. First, we recommend editing the following sentences *"Variation suggests that some entities can successfully mitigate a factor, as reflected by observed differences in outcomes, and that at least for a subset of entities is within their locus of control. It is important to remember that the locus of control should not be presumed to be fixed or immutable based on current practice, guidelines, and capabilities, or to be identical across diseases, conditions, entities, etc."* to more accurately reflect the degree of influence. We suggest replacing *"is within their locus of control"* with *"can be impacted"* as even successful organizations may not have 'control' but rather varying degrees of influence.
    b. We also recommend deleting the phrase *"It is important to remember"* to strike a more balanced tone in this section. We also recommend adding the following text: *"However, developers should also consider the impact that variation in the availability of community-based resources and state policy have on a measured entity's ability to mediate social risk. If mediation depends on the availability of certain resources, developers should consider if those resources are consistently available before removing a social variable from their model due to its ability to be mediated."*

Theme

Providing clarity

Sections Impacted

Conceptualizing the Model

Proposed Response

1. Thank you for your comment. The TEP acknowledged the purpose of NQF endorsement is to gain consensus among stakeholders about which measures warrant endorsement as the "best in class." This process requires the assessment of measure properties and testing against four endorsement criteria including importance to measure, scientifically acceptable, usable, and feasible. A key component of the scientific acceptability criteria is validity which evaluates the risk adjustment model of a candidate measure. Evaluation Committees examine an accountable entity's ability to influence a risk factor (i.e., locus of control), and this influences whether the risk factor should be included in the risk adjustment model or not. The Technical Guidance instructs developers to "*consider evidence that demonstrates that accountable entities can or cannot mitigate the effect of the social or functional risk factors linked to the measured outcome"* (pg. 30-31). In addition, the TEP and the Technical Guidance note that "*there may be statistical or practical reasons for considering a separate model for certain population subgroups… Developers may also consider excluding patients whose risk cannot adequately be predicted from the model or looking for more accurate risk adjusters for patients at highest risk (e.g., use of mechanical ventricular assist devices in patients with advanced heart failure to estimate their risk of mortality). Other reasons for a separate model might depend on the policy goals of the measure developer and measure implementer. The conceptual model can help developers to think about testing certain subgroups identified earlier in the development process. Developers should use caution when building separate models for subgroups unless there is sufficient sample size"* (pg. 38-39). Therefore, there may be opportunities to use different models depending on the program policy or other concerns.

Proposed Adjudication

2. On page 25, add **text:** *"Variation suggests that some entities can successfully mitigate a factor, as reflected by observed differences in outcomes, at least for a subset of entities,* ***those factors can be impacted*** (content has been edited out of the original). *The locus of control should not be presumed to be fixed or immutable based on current practices, guidelines, and capabilities, or to be identical across diseases, conditions, entities, etc.* ***However, developers should also consider the impact that variation in the availability of community-based resources and state policies have on a measured entity's ability to mediate social and/or functional risk. If mediation depends on the availability of certain resources, developers should consider if those resources are consistently available before removing a social or functional variable from their model due to its ability to be mediated.****"*

Questions for the TEP

- Does the TEP agree with the proposed responses?
- Does the Technical Guidance need additional changes to improve clarity?

Prompt 4: Please provide any other feedback on the technical guidance.

Comment #6

Commenter: Danielle Lloyd, America's Health Insurance Plan (AHIP)

Date Comment was Submitted: 9/21/2022

TEP Discussion Time Allotted: 15 minutes

Comment

1. AHIP thanks NQF and the TEP for the undertaking the difficult work to examine these challenging and nuanced issues. Measurement has become an important tool in our collective work to address health care disparities and promote health equity, especially in the context of the ongoing transition to value-based care. We must carefully balance the need to ensure high-quality for all with concerns about access and fairness.
2. As such, we recommend the executive summary take a more balanced approach to the reasons to adjust or not to adjust for social risk factors. NQF's 2014 guidance noted a belief that adjustment was essential for fair measurement in addition to preserving access for vulnerable populations and we recommend maintaining this consideration. We recommend adding language on page five after the text *"This potential unintended consequence of using unadjusted measures in value-based payment (VBP) programs could therefore reduce equitable access to care. Absent social risk adjustment, a provider caring for a higher proportion of socially disadvantaged patients might have a less favorable measure score if the difference in outcomes reflect social risks that quality care cannot fully address"* that states *"Supporters of adjusting for social risk factors also note that such adjustments are necessary to ensure fair comparisons and that measures truly reflect the performance of the accountable entity."*
3. The report also highlights the role of the program design and potential incentives in mitigating the effect of social risk factors in several places. While we agree VBC is potentially an important lever to promote equity, the use of a measure is outside the control of the NQF endorsement process. The Committees tasked with review are only able to assess if a measure is used, not how it is used. For example, the NQF process does not necessarily consider the use of a measure by health insurance providers or other private sector entities. Moreover, the design of private sector VBC and public reporting programs are not the purview of NQF. We recommend adding language clarifying that measures can be used by multiple parties and that program design may not be the same across measure users.
4. In addition, we ask that NQF edit the following sentence, *"Furthermore, properly structured payment incentives can ensure that providers caring for at-risk patients are not penalized."* This paper is about a consensus-based process for measure endorsement and should not assert that health plans are 'not properly' constructing their own incentive structures if they do not comport with NQF's stated ideal. Moreover, it should not be assumed that private payers can or will make up for a lack of risk adjustment on measures through risk adjustment within the payment structure. While this may be possible within federal programs, this may not be possible or even preferred among private plans. NQF should be structuring a system where measure developers can structure the measure as they see fit so long as it is sufficiently justified.
5. The report mentions in several places that risk adjustment decisions should be made in the context of the measure's use if known. For example, the Executive Summary notes that *"Developers should also examine the role of social and/or functional risk factors in the context of the measure's expected use, if known, as the measure's intended use may affect decision making regarding risk variable inclusion, feasibility of stratification, and potential unintended consequences."* Again, this is not always known, and measures can be used in multiple ways. We recommend adding language caveating that measures may have multiple uses and the same

measure could be used in different programs with different designs and incentive structures. During the trial period, if a measure was adjusted for social risk factors, developers were required to submit a version of their measure that did not include social risk factors in the risk model. The TEP could build on that previous guidance to allow the endorsement process to consider two version for a measure to account for differences in program design, rather than assume all implementers will use the same incentive or reporting scheme.

6. The report implies in several places a relationship between social and functional risk factors. The report states on page four that *"functional risk factors are important to examine since they may mediate the relationship between social risk and the measured health outcome."* The report also notes on page nine that *"Functional risk factors are important to examine since they may confound or mediate the relationship between social risk, quality outcomes, and resource use."* However, the report does not explain how functional risk could confound or mediate social risk. Moreover, the report seems to imply a one-way relationship: functional risk mediates or confounds social risk. However, the opposite could also hold true. Social risk could confound or mediate functional risk as a person's socioeconomic status could impact their ability to overcome limits in functional status. Clarification of the potential relationship and how these factors could mediate or confound each other could assist developers in developing their conceptual models to explore the potential pathways these risk factors could impact health outcomes.

Theme

Provide Clarity

Sections Impacted

Executive Summary
Conceptualizing the Model
Intended Use

Proposed Response

1. Thank you for your comments. We have provided responses within the "Proposed Adjudication" section below.
3. Thank you for your comments. We have provided responses within the "Proposed Adjudication" section below.

Proposed Adjudication

2. We thank you for your comment and suggest an addition to the language to achieve balance. The TEP should discuss the proposed edit as follows:
    a. *"This potential unintended consequence of using unadjusted measures in value-based payment (VBP) programs could therefore reduce equitable access to care. Absent social risk adjustment, a provider caring for a higher proportion of socially disadvantaged patients might have a less favorable measure score if the difference in outcomes reflect social risks that quality care cannot fully address.* ***Supporters of adjusting for social risk factors also note that such adjustments are necessary to ensure fair comparisons and that measures truly reflect the performance of the accountable entity"*** (pg. 5).
4. We thank you for your comment and suggest an addition to the language to achieve balance. The TEP should discuss the proposed edit as follows:
    b. *"Similarly, stakeholders acknowledge that determining whether measures should be adjusted for social risk factors will vary depending on each specific measure and the structure of the VBP or quality reporting program in which it is used.* ***A specific measure can be used by multiple parties and the program design (e.g., VBP models) may not be***

identical across measure users. *Stakeholders noted VBP models can be constructed to promote health equity by financing and rewarding providers for serving patients with social risk factors and by addressing their needs* (content has been edited out of the original). *Furthermore, payment incentives can ensure that providers caring for at-risk patients are not penalized* (content has been edited out of the original). *This can be achieved by making available to providers the resources they need to lessen the impact social risk factors have on outcome rates"* (pg. 5).

5. While this is mostly related to implementation of the Technical Guidance in NQF criteria, which is intended to be addressed and updated with these recommendations after this Technical Guidance if finalized, the TEP and the Technical Guidance note that "*there may be statistical or practical reasons for considering a separate model for certain population subgroups… Developers may also consider excluding patients whose risk cannot adequately be predicted from the model or looking for more accurate risk adjusters for patients at highest risk (e.g., use of mechanical ventricular assist devices in patients with advanced heart failure to estimate their risk of mortality). Other reasons for a separate model might depend on the policy goals of the measure developer and measure implementer. The conceptual model can help developers to think about testing certain subgroups identified earlier in the development process. Developers should use caution when building separate models for subgroups unless there is sufficient sample size"* (pg. 38-39)**.** Therefore, there may be opportunities to use different models depending on the program policy or other concerns. However, the TEP should discuss the below language addition to the intended use portion of the conceptual model.

   "The developer should explain, to the extent possible, how the intended use affects the developer's choices regarding risk variable inclusion, feasibility of stratification, and potential unintended consequences, ***recognizing that measures have different uses, and the same measure could be used in different programs with different designs and incentive structures"*** (pg. 30).

6. We thank you for your comment and suggest an addition to the language to achieve balance. The TEP should discuss the proposed edit as follows:
   a. *"However, there are strong differences of opinion as to whether adjusting for social risk factors achieves these goals. Furthermore, functional risk factors are important to examine since they may mediate the relationship between social risk and the measured health outcome.* ***In turn, social risk may confound or mediate functional risk as a person's social risks could impact their ability to overcome limits in functional status. [add citation Joynt Maddox, et. al., Health Affairs, Volume 38, No. 4, The Role Of Social, Cognitive, And Functional Risk Factors In Medicare Spending For Dual And Nondual Enrollees.]"*** (pg. 4)

Questions for the TEP

- Does the TEP agree with the proposed adjudications for balance?
- Does the Technical Guidance need additional changes to improve clarity?

Comment #7

Commenter: Doris Peter, Yale university School of Medicine

Date Comment was Submitted: 9/21/2022

TEP Discussion Time Allotted: 10 minutes

Comment

1. Thank you for the opportunity to comment, and thank you to the TEP members and NQF staff for their hard work on this project. In addition to the comment we made in response to the question about stratification, we have the following additional comments.
2. Conceptual models:
    a. The guidance asks developers to include the use of the measure in the conceptual model, but the conceptual model template does not invoke the measure's use. Could NQF provide an example within the templated conceptual model?
    The conceptual model is very complex; could NQF provide an actual example applied to an existing measure in addition to the template?
    Could NQF clarify what they mean in this statement on page 17; in particular the part of the sentence that we have highlighted in bold: *"More specifically, the conceptual model should delineate which factors can influence the health of the patient on their presentation for care versus the factors that might need to be considered in order to deliver truly patient-centered and effective care."*
3. When variables are not available to test in risk models:
    a. NQF asks developers to describe the direction of potential bias when risk variables are not available for testing. From page 20: *"The developer should describe the potential bias that may exist, and its direction, as a result of not including the risk factor(s) in the risk adjustment model."* This is complex. You could have variables that conceptually would have one direction in terms of potential bias, but when used with other variables in the model, they may end up working in the opposite direction due to interactions with other variables.
    b. NQF asks developers to justify why a measure is still valid even when a variable that is in the conceptual model cannot be tested. From page 32: *"If social and/or functional risk factor data are not available in a data source of sufficient quality (Table 1), but these factors were included in the conceptual model, the developer should describe the potential bias that may exist, and its direction, as a result of not including the risk factor(s) in the risk adjustment model. The developer should also provide a justification for why the measure still has validity in this circumstance."* Can NQF clarify what is acceptable in terms of justification? When deciding to adjust or not adjust for social/functional risk factors Page 26 states: *"When designing the conceptual model, it is important to remember that these factors can have either a direct or indirect effect (i.e., via the actions taken by the accountable entity) on the measured outcome. Both the direct and indirect effects of factors should be considered for model inclusion"* [page 26] We are concerned that the "indirect" effects may be accountable-entity characteristics (actions/behaviors or other) that you would not want to risk adjust as they may be related to the quality of care provided.
    c. The draft states that *"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."* We note that this ties the hands of developers and this should really be a balance between the conceptual and the empiric. If empiric results do not show a relationship between the variable and the outcome, then the developer could update their conceptual model

rather than include it in the risk model.

d. The current draft refers to methods of Integrated discrimination improvement (IDI) and the net reclassification improvement (NRI) approaches. We note that NQF may want to state that these methods should be used judiciously. Please see this reference: https://hbiostat.org/blog/post/addvalue/ "Especially problematic are measures such as the categorical version of NRI (net reclassification improvement) which not only requires arbitrary categorization of risk estimates but then goes on to use inefficient binary summaries from them. Pencina (personal communication) has regretted including statistical tests for these measures in his highly-cited paper, as these tests have nowhere near the power of the gold-standard likelihood ratio test."

4. Calibration
   a. The draft guidance states that *"Measures that include social or functional risk factors in the final risk model should be calibrated in subgroups defined by those factors to the extent possible"* We recommend that this requirement be for all measures, not just those that include social or functional risk factors. Appendix D [page 77] Measure 1789 is attributed to the wrong developer. It was developed by Yale/CORE.

Theme

Developer burden
Provide Clarity

Sections Impacted

Technical Guidance Overview
Conceptualizing the Model
Appendix D: Examples of Approaches to Social and/or Functional Risk Adjustment

Proposed Response

1. The intention is for the graphical depiction of the conceptual model to be accompanied by a narrative description, including discussion of intended use. As in the Technical Guidance: *"…the design of the conceptual model …should be outlined in a narrative description of the model as well. This will help others who are uninvolved in the conceptual model's development to understand what decisions were made and why"* (pg. 25). Also, *"Some of these methods have been added as illustrative examples from NQF-endorsed measures (Appendix D)"* (pg. 17). Please refer to Appendix D for an example of how the conceptual model is applied in a measure submission. Additionally, the measures provided in Appendix D are intended to be illustrative examples and are used to support the best practice standards, where appropriate. Since the Technical Guidance standards have not yet been implemented into NQF criteria at this stage, Appendix D will not include illustrative measure examples for every aspect of the Technical Guidance recommendations.
2. The statement from page 17, "*More specifically, the conceptual model should delineate which factors can influence the health of the patient on their presentation for care versus the factors that might need to be considered in order to deliver truly patient-centered and effective care"* is to clarify that although patient characteristics at the start of care are not fixed, accountable entities may be able to take actions that are responsive to those characteristics in order to deliver overall improved quality care.
3. Regarding acceptable justifications for why a measure still has validity despite a lack of a specific factor, the TEP has stated that a literature review and description of the available evidence should be used to justify a conceptual model. If research does not exist in an area, a narrative

description of the conceptual model allows for the opportunity to highlight gaps or areas where new research is needed.

4. The TEP acknowledges that there will always be the possibility of missing data resulting in an imperfect model. As a result, the TEP suggests that the developer review the literature and determine how a risk factor might affect particular subsets of the measured entity based on how those patients at-risk are distributed across those providers.

Proposed Adjudication

None.

Questions for the TEP

- Does the TEP agree with the proposed responses?
- Does the Technical Guidance need additional changes to improve clarity?
- The commenter noted that methods by Pencina should be "used judiciously." The TEP should discuss whether this statement should have a caveat added.

Comment #8

Commenter: Joyce Lam, Acumen, LLC

Date Comment was Submitted: 9/21/2022

TEP Discussion Time Allotted: 10 minutes

Comment

1. Acumen, LLC appreciates the opportunity to provide comments on the Draft Report. While we appreciate NQF's efforts to develop a risk adjustment framework, we continue to have concerns that this report fails to provide clarity about the actual standards that should guide risk adjustment decisions or evaluation, and instead risks increasing burden, subjectivity, and inconsistency in the evaluation process.
2. First, this report omits discussion of how developers or evaluation panels should determine the best course of action to take when examining testing results. That is, measure developers/owners and evaluation panels need guidance for what sorts of results indicate the need to adjust for SRFs and what testing results should translate to a pass or fail rating on validity. Without these standards being addressed, the guidance risks exacerbating the existing concerns about inconsistency and unreliability in measure evaluation.
3. Second, the guidance does not elucidate how to distinguish between the two conceptual channels for disparities that may be present in testing results, namely that reasons for disparity can be within or outside of the provider's locus of control. Risk adjustment should only occur if the latter is the dominant explanation. However, the empirical testing described in the draft report do not allow one to distinguish between these two cases. We believe that this is critical to address before these standards can proceed towards implementation.
4. Finally, we are concerned that the guidance creates substantial burden and procedural uncertainty without commensurate benefits to the rigor or consistency of approaching the question of whether to adjust for SRFs. In its current form, the report serves as a list of what minimum information evaluation panels require without addressing practical matters. For instance, there are hundreds of SRFs that a developer could research and test. This leaves developers in the position of either conducting a prohibitive amount of work or guessing at an evaluation panel's preferences with more limited testing. NQF's agnosticism to measure use also unintentionally curtails the utility of an endorsement evaluation.
    a. For instance, a developer might be able to provide specifications that are stratified to

enable scores to be calculated and displayed for at-risk subgroups, but a measure reporting contractor would face many different constraints for the calculation and display of measures.

5. We thank NQF for the opportunity to comment, and urge them to keep in mind the aims of CDP redesign discussed in April 2022 when considering next steps for this guidance. In particular, we note that the aims included strengthening consistency and science oversight, lowering burden for developers and volunteer committee members, and increasing overall efficiencies.

Theme

Developer burden
Provide Clarity

Sections Impacted

Conceptualizing the Model
Empirically Testing Risk Factors

Proposed Response

1. The TEP is not prescriptive on what testing results must show and what testing results should translate to a pass or fail rating on validity for several reasons.
    a. First, the TEP reemphasizes the centrality of the conceptual model in determining whether factors should be included in the final model and the informative but not deterministic role of empirical testing in selecting a social risk factor for risk adjustment.
    b. Second, it is dependent on the outcome being measured.
    c. Third, the TEP wants to allow for flexibility in measure development and measure innovation.
    d. Finally, this requires input from NQF governing bodies. This concern is related to the implementation of the Technical Guidance in NQF criteria, which is intended to be addressed and updated with these recommendations after the Technical Guidance is finalized. Furthermore, the Technical Guidance acknowledges this future work on page 44, stating that *"While this project gathered input through an environmental scan and from a multistakeholder TEP, focus groups, and public comment, specific changes to NQF's CDP [consensus development process] require several important steps prior to the implementation of this guidance."* The Technical Guidance also outlines these future steps and the overall implementation approach.
3. The Technical Guidance offers developers in-depth instructions to produce thorough defenses of the conceptual model, upon which further decisions regarding risk adjustment depend. The TEP reiterates the centrality of the conceptual model and its relationship to final decisions for including or not including SRFs. Again, the TEP wants to allow for flexibility in measure development and measure innovation and these decisions are dependent on the outcome being measured.
4. Conceptual models should be formed based on the literature to contribute to the rigor of decisions regarding whether to adjust for SRFs. Factors in Minimum Standard #2 should supplement what is identified through the literature search conducted in Minimum Standard #1. These minimum standards focus on selection of factors for consideration in the conceptual model and are not required for inclusion in either the final conceptual model or for the testing or final implementation of the risk adjustment model. NQF acknowledges that we do not oversee implementation of measures or their reporting.

Proposed Adjudication

2. The Technical Guidance does not outline how to distinguish between the two conceptual channels for disparities empirically. Rather, it points out that "*Developers can demonstrate an accountable entity's ability to meaningfully influence a factor by citing the primary literature, public reports, and case studies, and/or by conducting empirical analyses…[For example,] variation suggests that some entities can successfully mitigate a factor, as reflected by observed differences in outcomes, and at least for a subset of entities, factors* ***can be impacted…However, developers should also consider the impact that variation in the availability of community-based resources and state policies have on a measured entity's ability to mediate social risk. If mediation depends on the availability of certain resources, developers should consider if those resources are consistently available before removing a social or functional variable from their model due to its ability to be mediated…****evidence to support these decisions can…be taken from a combination of sources, such as expert opinions, literature review of peer-reviewed articles and white papers, and/or from conducting internal empirical analyses…Furthermore, the conceptual model should consider whether it is feasible for accountable entities to diminish the impact of social or functional risk factors, as measurement should be accurate and reflect the agreed-upon scope of responsibility of the accountable entity, as opposed to being too aspirational in expanding the locus of control of the measured entity beyond its achievable scope. Developers should consider evidence that demonstrates that accountable entities can or cannot mitigate the effect of the social or functional risk factors linked to the measured outcome."* (pgs. 24-30) As stated above, the Technical Guidance is not prescriptive about which results must be relied upon. Again, the TEP should review and discuss how to use these empirical results to add guidance.

Questions for the TEP

- The TEP should review and discuss how to use empirical results to add guidance.
- Does the TEP have any recommendations on addressing constraints that the commenter has raised related to reporting and how the Technical Guidance can be updated to address the concern?
- Does the TEP agree with the proposed responses?
- Does the Technical Guidance need additional changes to improve clarity?

Comment #9
Commenter: Lisa McGonigal, Kidney Care Partners

Date Comment was Submitted: 9/20/2022

TEP Discussion Time Allotted: 10 minutes

Comment

1. Kidney Care Partners (KCP) appreciates the opportunity to comment on NQF's draft technical guidance report, Developing and Testing Risk Adjustment Models for Social and Functional Status-Related Risk within Healthcare Performance Measurement. KCP is a coalition of 32 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.
2. KCP has long supported efforts to assess and account for social risk factors in performance measurement through adjusters and other mechanisms. As a matter of policy, we believe all measures used in federal accountability programs should be examined 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 providers, promote health equity, and empower patients and their care partners. As such, KCP appreciates the thoughtful recommendations put forth in this report. We agree with the Technical Expert Panel 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 differences across accountable entities in 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 performance measures and value-based payment programs. Nevertheless, we have a number of concerns with the TEP's guidance in the report.
3. First, we again highlight that the recommendations remain 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 has now clearly outlined the different points of view about whether adjusting for social risk factors that influence measured outcomes will accelerate or hinder efforts to improve health equity, including ASPE's stance that adjusting outcome measures for social risk may obscure real differences in care. However, despite the acknowledgement, no remedy or rationale for this discrepancy is offered in the report, and we remain concerned that such a striking inconsistency risks perpetuating ambiguity and confusion among measure developers and other stakeholders.
    a. References:
    [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.
4. We also appreciate the TEP's clarification regarding the use of risk stratification as 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. It is explicitly stated within the report that, generally speaking, neither process measures nor "never

event" outcome measures should be risk adjusted. However, we again note that there are many outcome measures for which it may be found that there is no clinical rationale to support social risk adjustment; it is in such instances that risk adjustment may in fact obfuscate real differences in performance and paradoxically reinforce health inequities. 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 many outcome measures (i.e., not just "never events"), 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.

   a. References:
      [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).

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 some detail what types of testing methodologies are—and are not—acceptable.
   a. For example: *"Simple bivariate and multivariable tests alone should not determine whether a social or functional risk factor is included in the risk model" . . . "Changes in model discrimination, such as c-statistics, can inform but should not determine decisions on whether to include an additional social and/or functional risk factor in the model specification."*
6. 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,
   a. [1] 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."*
   b. [2] 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.
      References:
      [1] National Quality Forum. Social Risk Trial Final Report. July 2017.
      [2] National Quality Forum. Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors Technical Report. August 2014.

Theme

Developer burden
Provide Clarity

Sections Impacted

Core Principles

Proposed Response

3. The TEP acknowledged the ASPE report in the base period. Similar to the ASPE report, this Technical Guidance document acknowledges the importance of data collection and interoperability standards for social and functional risk. 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. It recommends that risk adjustment be based primarily on the conceptual model. The TEP acknowledges ASPE's important work in this space, but to clarify, this Technical Guidance is intended to focus on risk adjustment for the purposes of NQF measure endorsement.
4. Stratification can be used in conjunction with or as an alternative to risk adjustment, depending on the conceptual model and all outcome measures require the same level of consideration when making decisions about risk adjustment and stratification. The TEP acknowledges that the appropriateness of social risk adjustment for outcome measures depends on the conceptual model as described in the report as follows: "*…the purpose of risk adjustment in measurement is to create measures that fairly compare quality of care between measured entities. Stratification enhances transparency to highlight known areas where disparities exist, which would be researched as part of the construction of the conceptual model. This minimum set of variables guides developers to focus on patient groups with known, widespread disparities. For this subgroup analysis and for the groups identified in the conceptual model, a measure developer's capabilities may also depend on available data and sample sizes within subgroups. If this is the case, measure developers should demonstrate why the analysis was not feasible*" (pg. 41).
5. The TEP acknowledges the comment regarding measure developer burden and reiterates the centrality of the conceptual model and its relationship to final decisions for including or not including SRFs. The TEP also acknowledges that bivariate and multivariate tests are informative but not individually deterministic for inclusion or not of social risk factors. The approach outlined in the technical guidance requires developers to not only determine empirically but also conceptually "*the hypothesized pathways between the social and/or functional risk factors, patient clinical factors, healthcare processes, and the measured healthcare outcome. The pathways between risk factors and the care process should be illustrated and accompanied by evidence of the relationships"* (pg. 21).

Proposed Adjudication

None.

Questions for the TEP

- Does the TEP agree with the proposed responses?
- Does the Technical Guidance need additional changes to improve clarity?