NATIONAL QUALITY FORUM
Driving measurable health improvements together

Meeting Summary

Best Practices for Developing and Testing Risk Adjustment Models, Technical Expert Panel – Web Meeting 4 Public Comments Review

The National Quality Forum (NQF) convened a public web meeting for the Best Practices for Developing and Testing Risk Adjustment Models Technical Expert Panel (TEP) on October 24, 2022.

Welcome, Roll Call, and Review of Web Meeting Objectives

Matt Pickering, NQF senior director, and the TEP co-chairs, Karen Joynt-Maddox and Philip Alberti, provided welcoming remarks to the participants. Hannah Ingber, NQF manager, facilitated roll call and reviewed the meeting objectives, which were to discuss and adjudicate public comments received during the public commenting period and to finalize the Technical Guidance.

Review and Discuss Public Comments on Technical Guidance

Prior to the web meeting, the TEP was sent a copy of a memo that included the public comments received on the updated Technical Guidance. Also included in the memo for the TEP's consideration were NQF-drafted proposed responses to the comments and text edits to the Technical Guidance.

During the web meeting, Dr. Pickering summarized the comments and requested that the TEP should express agreement or disagreement with the proposed responses and the suggested text edits to the guidance. Dr. Pickering stated that nine comments were received from four organizations, three of which are measure developers. He noted that several comments were submitted by Danielle Lloyd on behalf of America's Health Insurance Plans (AHIP). Therefore, since Danielle Lloyd is currently on the TEP, Dr. Pickering stated that she will be recused from the discussion of AHIP's comments due to a perceived conflict of interest. However, Ms. Lloyd will be available to answer any questions the TEP may have with respect to AHIP's comments.

Dr. Pickering stated that NQF received comments for each of the following prompts: feedback on the recommendations for stratification, feedback on whether the minimum standards need additional specification or clarifications, feedback as to whether there are challenges to operationalizing these minimum standards, and lastly, any general feedback on the Technical Guidance. He further noted that there were common themes amongst the comments, which focused on providing more clarity, burden to developers, and requirements for NQF's Consensus Development Process (CDP) (i.e., endorsement).

The memo document was displayed on the Zoom platform to facilitate the TEP's discussion, and a summary of the TEP's deliberations is provided below.

Prompt 1: Feedback on the Recommendations for Stratification

For this first prompt, Dr. Pickering summarized that a comment was received from Yale University, School of Medicine, which requested more clarification as to the application of stratification and noted concern with the added measure developer burden due to the stratification requirements. For the proposed response, Dr. Pickering noted that the intent of stratification is to enhance transparency by showcasing where disparities exist, and that stratification should be considered independently of

decisions regarding risk adjustment. Furthermore, Dr. Pickering emphasized that the TEP recognizes that stratification may add requirements to the NQF endorsement process for developers, but the added requirements will advance uncovering disparities and further promote health equity. Furthermore, the guidance does recognize certain limitations, such as sample sizes and data availability, which can impact the ability to perform stratification. Dr. Pickering noted that based on these comments, no proposed adjudications were made to the text and proceeded to open the discussion by asking the TEP if there was general agreement with the proposed responses or if additional changes were needed to the guidance.

The TEP agreed with the proposed responses and recommended that there be a transition period for when these guidance recommendations take effect within NQF criteria so that developers have time to plan and prepare for the new endorsement requirements. Dr. Pickering responded, noting that the guidance does outline the additional work needed to incorporate the recommendations into NQF criteria, which includes providing education and training to measure developers and other NQF stakeholders. A TEP member also suggested adding a note early on in the Technical Guidance about what this guidance is intended to do and what it is not intended to do. Specifically, this guidance is not intended to be a checklist of things, but a guidance around best practices. There was no disagreement from other members of the TEP for this suggested addition.

Moving to the second comment under this prompt, Dr. Pickering noted that the comment was from Acumen LLC, which requested clarification as to whether the stratified reporting is intended to be done prior to or after risk adjusting for social risk factors and clarification as to the purpose of stratification analyses, including what the implications are for the [NQF] evaluation process, the requirements for reporting stratified results, and the implications of adjusting or not adjusting when within entity variation is zero for some and non-zero for others.

In summarizing the proposed responses, Dr. Pickering stated that stratified reporting is intended to be done prior to risk adjustment for social and/or functional risk by using either unadjusted or clinically risk-adjusted data, as noted in the guidance. He further stated that stratified reporting is a separate issue, not one that always precedes a risk adjustment approach. It may sometimes be performed in situations where risk adjustment is not planned, and in other scenarios it may be performed in parallel, not in series, with a risk adjustment approach. If measure scores differ across subpopulations that have high and low social risk factors, measure developers should acknowledge this difference in measure scores, but this difference should not determine whether social risk factors be included in the model or not. With respect to requiring the reporting of stratified results, Dr. Pickering noted that NQF and the TEP recognize that neither are able to specify the requirements for public reporting of measure stratification results. Lastly, Dr. Pickering noted that no proposed edits were made to the guidance based on these comments and proceeded to open the discussion by asking the TEP if there were any concerns with the proposed responses and what the implications are for adjusting or not adjusting when within entity variation is zero for some and non-zero for others?

A TEP member responded stating that there are no implications of zero or limited within disparities of measure validity or risk adjustment, specifically noting this can mean superior or low performance by a provider due to small disparities. Dr. Pickering probed the TEP by asking if any text changes were needed. The TEP agreed that it is not able to require public reporting requirements of stratified results but noted that requiring stratification specifications for endorsement enables the ability to publicly report those results. The TEP agreed no changes were required, and Dr. Pickering concluded the review of prompt one and moved to the next prompt.

Prompt 2: Feedback on Whether the Minimum Standards Need Additional Specification or Clarifications

For this prompt, Dr. Pickering proceeded to summarize the first series of comments from Acumen LLC, which raised concern with minimum standard #1 (requiring the development of a conceptual model) conflicting with minimum standard #2 (requiring a minimum set of social and functional factors that should be considered in the conceptual model). In addition, Acumen raised concern with the added burden of minimum standard #1, due to the hundreds of possible risk factors that could theoretically have a relationship with the outcome of interest.

For the proposed response, Dr. Pickering stated that minimum standard #2 instructs the developer to consider a set of factors for the conceptual model, which are not required for inclusion in either the final conceptual model or the final risk adjustment model. The factors in minimum standard #2 should supplement what is identified through the literature search conducted in minimum standard #1. He further noted that the Technical Guidance acknowledges the burden to developers by not being overly prescriptive. However, at a minimum, the developer should conduct a literature review to inform the conceptual model, which can be supplemented with the grey literature and expert opinion. Additionally, the literature review is not required to be a systematic review either. Lastly, in response to this comment, Dr. Pickering proposed a text edit to minimum standard #2 language to clarify that consideration of these factors is for their "potential inclusion in the conceptual model before it is finalized." Turning to the TEP for discussion, the TEP did not have any concerns with the proposed responses, or the proposed text edit to minimum standard #2.

Dr. Pickering proceeded to summarize Acumen's comments regarding minimum standard #3 (the developer should describe the potential bias that may exist if a risk factor is in the conceptual model, but not available in a data source). The comment noted that this standard appears to set the presumption of measure invalidity, and that this shifts the burden of proof to the developer to show that a measure is valid when there is no data for potential risk factor variables. For the proposed response, Dr. Pickering confirmed that it is the developer's responsibility to analyze the bias impact of a variable that was identified as important in the conceptual model, but unavailable in a data source. Additionally, the Technical Guidance states that at a minimum, developers should examine previously published evidence to help estimate the directionality of the bias for the factor of interest. No text edits have been proposed and the TEP did not raise any concerns with the proposed response.

Lastly, Dr. Pickering noted that Acumen's last comment was with respect to minimum standard #5 (provide descriptive statistics on how the risk variables identified from the conceptual model are distributed across the measured entities). The comment raised concern that this requirement adds burden to both developers and evaluation panels and that by having more guidance as to what risk factors should be included in minimum standard #1, the burden associated with minimum standard #5 will be contained. For the proposed response, Dr. Pickering stated the intent of minimum standard #5 is to better understand the populations served across accountable entities and the potential disparities. He noted the proposed addition to the guidance, which was to note that "if there is a high degree of difference in social risk 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."

Opening up the floor for TEP discussion, the TEP suggested revising the text edit by splitting the sentence, as excluding accountable entities and the use of proxy variables are not a solution to the same problem. A TEP member questioned if the intent is to exclude accountable entities simply based on not having risk factor data. Another TEP member commented that the intent was to exclude accountable

entity outliers, not exclude an accountable entity due to the lack of data. Co-chair. Joynt-Maddox suggested that NQF revise this edit to separate out these two concepts. Dr. Pickering stated that NQF will make the requested revisions and will seek the TEP's feedback outside of the web meeting.

Prompt 3: Feedback as to Whether there are Challenges to Operationalizing these Minimum Standards

For the third prompt, Dr. Pickering summarized the first set of comments, which were from Acumen LLC. The comments suggested clarifying the intent of the minimum standard #2, namely how developers can demonstrate that they have fulfilled this minimum standard and whether the intent is to use gender or biological sex (or race or ethnicity), how these are captured from data sources, their relationship with a health outcome, and the implication if they are used in risk adjustment.

For the proposed response to this comment, Dr. Pickering highlighted that the TEP's intent was not to be overly prescriptive as to whether gender or biological sex is appropriate, as it may depend on the outcome being measured and the data available. Furthermore, 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. Dr. Pickering noted that no guidance edits are proposed in response to this comment. The TEP did not raise any concerns with the proposed response.

Dr. Pickering then summarized Acumen's comment for minimum standard #6 (model calibration should be conducted within the overall population and within relevant subgroups), which stated that the guidance appears to imply that none of the three empirical tests (predictive ability, discrimination, and calibration) are determinative in considering whether to risk adjust for SRFs and functional risk, but that this standard is requiring that developers conduct these empirical tests. Lastly, there is no guidance for evaluation panels about how they should use these empirical results.

For the proposed response, Dr. Pickering summarized that empirically testing the associations of risk factor to the measured outcome is not deterministic for inclusion into the model. Rather, a developer may consider these tests for decision making by reporting on any associations in the literature, supplemented by expert opinion, and the magnitude of the results. With respect to empirically testing risk model performance, such as calibration analyses, this should be assessed with and without social risk factors for relevant subpopulations. However, Dr. Pickering noted that the TEP should review and discuss how to use these empirical results for measure evaluation.

Circling back on the previous TEP suggestion to a comment in prompt #1, co-chair Alberti iterated that the purpose of the guidance is not to be a checklist and that the guidance would need evolve over time. To help developers use the guidance, a TEP member suggested including a "why this is important" statement to support the five major steps in the Technical Guidance. Other TEP members agreed that it would be beneficial to clearly state the importance of each step in the respective sections. Dr. Pickering stated that the project team will incorporate these importance statements for each step.

Moving to the last comment for prompt #3, Dr. Pickering noted that this was submitted by Danielle Lloyd on behalf of AHIP. He summarized that the comment expressed that mediators of social risk factors 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. Therefore, this variation may limit or enhance a measured entity's ability to mediate the impact of potential social risks on health outcomes.

Dr. Pickering stated that AHIP proposed a suggested edit to the guidance for the TEP's consideration, which notes that "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 mediate."

During the TEP discussion of this comment and proposed edit, co-chair Alberti suggested noting in the conceptual model section that developers consider the variability in the availability of community-based resources and state policies. Another TEP member suggested including the consideration of provider availability as well. Another TEP member emphasized that some models are dependent on available resources in the community, and it is difficult to quantify all of the things that impact the end points of measurement. The same TEP member suggested that the developer should state this in the discussion of their model as a limitation.

Dr. Pickering stated these suggestions will be added to the guidance. Other TEP members did not raise and concerns with the proposed suggestions, and no further comments were provided by the TEP.

Prompt 4: Other Feedback on the Technical Guidance

For the last prompt, Dr. Pickering noted that the first comment was also submitted by AHIP, which suggested that the executive summary take a more balanced approach to the reasons to adjust or not to adjust for social risk factors. AHIP proposed adding the statement, "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." Additionally, AHIP noted that there is variability in program design and that the evaluation bodies are only able to assess if a measure is used, not how it is used. Therefore, AHIP suggested noting that "a specific measure can be used by multiple parties and the program design (e.g., VBP models) may not be identical across measure users." AHIP further emphasized that the guidance 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. Lastly, AHIP noted that the guidance implies that there is a mediating relationship between social and functional risk factors, but it does not explain how functional risk could be confounded or mediated social risk.

Before turning the TEP for discussion, Dr. Pickering asked whether the TEP agrees with AHIP's suggested edits. With respect the confounding relationship that may exist with functional and social risk, the TEP suggested adding an example of how social factors mediate functional status.

A TEP member expressed concern with the proposed edit that indicates that a measure could be used in different programs with different designs and incentive structures. The TEP member commented that a measure should not be designed to ensure that it can be used in any program. Rather, the developer should acknowledge the trade-offs of a measure and its risk adjustment approach that are designed for a specific program context and that its use in another program may require a different approach. The TEP member further stated that this language already exists in the conceptual model section of the guidance.

Regarding the language "properly structured payment incentives", a TEP member suggested editing slightly to state, "furthermore, incentive payment structures can help ensure providers caring for at-risk patients are not penalized." The TEP did not raise any additional concerns to the other proposed edits from AHIP.

Dr. Pickering then summarized a comment submitted from Yale University School of Medicine, which asked for an example of a measure that included the measure's use in the conceptual model. The comment also asked for clarification on the statement, "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." Yale also sought clarification as to what is acceptable in terms of justification of a measure's validity even when a variable that is in the conceptual model cannot be tested.

In response to these concerns, Dr. Pickering stated that Appendix D of the guidance provides an example of how the conceptual model is applied in a measure submission, and that the measures provided in Appendix D are intended to be illustrative examples to support the best practice standards, where appropriate. However, 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. With respect to the delineation statement, Dr. Pickering clarified 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. Lastly, the TEP acknowledges that there will always be the possibility of missing data resulting in an imperfect model. As a result, the TEP suggests in the guidance 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. Dr. Pickering noted that there were no proposed edits for the TEP's consideration for these comments. The TEP did not raise any concerns with these proposed responses.

Dr. Pickering continue to note other concerns identified in Yale's comment, namely that the current draft refers to methods of integrated discrimination improvement (IDI) and the net reclassification improvement (NRI) approaches. Yale suggested that the guidance state that these methods should be used judiciously. The TEP agreed with this addition to the text where these methods are mentioned.

Lastly, Yale's comment noted that 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." Yale recommend that this requirement be for all measures, not just those that include social or functional risk factors. The TEP agreed with this proposed edit.

Dr. Pickering moved the next comment, which was from Acumen, LLC. The comment raised concern that the report omits discussion of how developers or evaluation panels should determine the best course of action to take when examining testing results. In response to this comment, Dr. Pickering iterated that the TEP is not prescriptive on what the testing results must show and what testing results should translate to a pass or fail rating on validity as this concern is related to the implementation of the Technical Guidance in NQF criteria, which is intended to be addressed after the Technical Guidance is finalized. The Technical Guidance also outlines these future steps and the overall implementation approach. The TEP did not raise any concerns with the proposed response.

Lastly, Acumen raised concern that the guidance does not elucidate how to distinguish between the reasons for disparity, which can be within or outside of the provider's locus of control, and that in its current form, the guidance serves as a list of what minimum information evaluation panels require without addressing practical matters. Dr. Pickering summarized that the edit made from prompt #3 will address the disparity concern. He reminded the TEP that this edit states that "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." In response to Acumen's last concern, he stated that 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 did not raise any concerns with these suggestions and responses.

For the last comment, Dr. Pickering noted that Kidney Care Partners submitted it. The commented noted that the recommendations in the guidance remain at odds with the Office of the Assistant Secretary for Planning and Evaluation's (ASPE) advisory reports with no remedy or rationale for this discrepancy is offered in the report. The comment also mentioned that there are many outcome measures for which it may be found that there is no clinical rationale to support social risk adjustment and that NQF allow developers discretion in this regard. Lastly, the comment raised concern that these are rigid recommendations, particularly for small developers.

Dr. Pickering summarized that similar to the ASPE report, this Technical Guidance document acknowledges the importance of setting 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. 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. With respect to risk adjustment with no clinical rationale, Dr. Pickering stated that the TEP acknowledges that the appropriateness of social risk adjustment for outcome measures depends on the conceptual model, which is left up to the discretion of the developer. In response to the commenters last concern, Dr. Pickering noted that 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 social and/or functional risk. The TEP did not raise any concerns with these responses.

Public Comment

Dr. Pickering opened the web meeting for public comment. No public comments were offered.

Next Steps

Simone Bernateau, NQF analyst, summarized that a final summary of the meeting discussion will be posted on the project webpage on November 16. Ms. Bernateau expressed that the TEP feedback will be incorporated into the Technical Guidance after today's meeting, and the final Technical Guidance will be posted on the project webpage on December 21, 2022.

Dr. Pickering noted for the additional TEP-requested revisions that were discussed today, the project team will be sharing these with the TEP after the web meeting for review. He then thanked the TEP, including the leadership of its co-chairs, Karen Joynt Maddox and Philip Alberti, the Federal Liaisons, and the members of the public for their time and participation. He then adjourned the call.