



Best Practices for Developing and Testing Risk Adjustment Models, Technical Expert Panel – Web Meeting 3

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 July 13, 2022.

Welcome, Roll Call, and Review of Web Meeting Objectives

Matt Pickering, NQF senior director, and the TEP co-chair, Karen Joynt Maddox, provided welcoming remarks to the participants. Hannah Ingber, NQF manager, facilitated roll call and reviewed the meeting agenda and objective, which was to obtain TEP-input on updates made to the base period [Technical Guidance](#), with a focus on specific elements that need further clarification or consensus.

Review and Discuss Technical Guidance Updates

During this portion of the web meeting, Dr. Pickering shared the main areas within the Technical Guidance that underwent a series of modifications. These included, specifications and requirements (i.e., expectations for developers) for both stratification and for risk model calibration, key areas of consensus, including how the consideration of race is presented within the guidance, and other key Technical Guidance updates, which were derived from a review of stakeholder feedback during [web meeting #2](#). The key sections on the updated Technical Guidance were displayed on the Webex platform along with questions for the TEP, which the co-chair utilized to facilitate the discussion.

Due to time constraints, the TEP was not able to discuss all topics noted above. For the topics that were discussed, a summary of those deliberations is provided below. For each topic discussed, the TEP provided feedback on the updates within the Technical Guidance, including adding clarification to some of the standards and recommendations for the stratification and calibration sections.

Stratification Specifications and Expectations

Dr. Pickering explained that stratification in the Technical Guidance refers to the division of a population or resource service into distinct, independent strata or groups of similar strata that enables analysis of the specific subgroups. As mentioned in the guidance, measures that adjust for social or functional risk factors, such as dual eligibility and disability status, generally make disparities in care for patients with the risk factor less visible. Therefore, measures that include social or functional risk adjustment in the final measure specification should, at a minimum, also assess a stratification approach for displaying measure scores by at-risk subpopulations. Dr. Pickering noted that this is reflected in minimum standard #7 within the guidance. He stated that for the purposes of this guidance, stratification specifications for NQF consideration should address how the results are calculated and the reporting approach. Yet, stakeholders requested more guidance on what the approach and specifications should be for stratification, more guidance on determining what variables should be stratified, and whether this can be done together (i.e., adjusting for and stratifying by the same risk variable), as participants disagreed and stated that this is not appropriate.

Co-chair, Dr. Maddox, led the discussion by asking the TEP to determine the “who” (i.e., what populations or variables should developers stratify by) and the “how” (i.e., what stratification strategies should developers use) with respect to stratification. A TEP member suggested that NQF develop a list of key patient characteristics (i.e., social and functional risk factors) that are of general interest for quality measures, excluding those characteristics in which there are limited data. The same TEP member suggested that the list could include sexual orientation and gender identity. Another TEP member suggested being broadly inclusive in developing a list of characteristics, expressing concerns about being too prescriptive and that each developer may not have access to certain data sets/data sources. Dr. Maddox shared that Medicare data are easily available within various data sets and asked if there were other data source considerations. A TEP member shared that beneficiary eligibility information is a good data element to use in identifying the population. Additionally, the same TEP member suggested building a database for disability status, since it’s related to patient eligibility for certain healthcare programs, such as dual eligibility for Medicare and Medicaid.

Rather than be overly prescriptive, another TEP member suggested creating a list of variable categories for stratification analyses, so that developers may be able to identify a specific variable within the data set the developer is using to develop and test the measure, which also aligns with the specified category. For example, a variable category could be an “indicator of disability status”, in which the variable of dual eligibility status may be used. Another TEP member expressed that instead of creating a list, there could be a menu of characteristics or categories that is suggested for developers to consider for stratification. The developer can then determine what variables exist in their data that align with those categories. For example, there could be a category of geographic location, which the developer may consider using variables such as census tract, zip code, region, etc. for stratification, depending on their respective data availability. Another TEP member added that if there are data availability concerns, the developer should provide a rationale for their set of factors for stratification. This rationale would include why they didn’t include a variable for a specific menu category, which may be due to data resource constraints and/or data collection burden.

A TEP member noted that perhaps the characteristics used for stratification align with the minimum set of factors that have been identified for consideration within the conceptual model. The TEP agreed that some, but not all of those factors should be required, at a minimum for stratification. Therefore, similar to the approach to variable examination in the conceptual model, the TEP recommended that a minimum set of variables be used to test for subgroup stratification analysis, including disability status, ethnicity, race, and rurality. Beyond this minimum set of variables, developers should consider stratification to distinguish between groups of patients who may have difficulties accessing care, for example, as suggested in the literature, by patients, by experts or by other stakeholders, and as reflected in the conceptual model. These variables should influence the outcome but themselves do not reflect actions by the accountable entity (i.e., are not related to the accountable entity’s care delivery behaviors).

A TEP member asked for clarity on the term “report on” with respect to stratification and if it was for NQF endorsement or for public reporting. Dr. Pickering responded that for the purpose of the Technical Guidance, the reporting for stratification would be for NQF endorsement. The same TEP member suggested including this distinction within the Technical Guidance. One TEP member asked as part of the NQF’s endorsement maintenance for measures, will developers be required to report on the stratification approaches being recommended within the guidance? Dr. Pickering responded that if these recommendations are included in NQF’s measure endorsement criteria, developers will be required to report the stratification approaches in maintenance evaluations of measures. Additionally,

Dr. Elizabeth Drye, NQF's chief scientific officer, shared that the modification of the endorsement criteria process will be done through engagement with stakeholders, including members of the public.

Dr. Maddox then moved to the "how" and asked the TEP members to share the methodological limitations to stratification and/or best practices that should be considered for measure developers. A TEP member shared that it's possible to present adjusted, stratified results and suggested providing concrete examples. The TEP expressed that stratification of measure results can be deployed regardless of whether a measure has been risk adjusted for social and/or functional risk adjustment. Stratification and risk adjustment for social and/or functional risk can be jointly deployed, or a measure developer may determine that stratification alone may be most appropriate. In the case that both stratification and risk adjustment for the social and/or functional risk factor are used, the social and/or functional risk factor used for the stratification analyses is not included in the risk adjusted calculation of the stratified results by subgroup. The TEP discussed that stratification can be implemented by displaying raw rates for the overall population and between subgroups or by displaying risk-adjusted rates (without the focal social and/or functional risk factor) depending on the measure focus and intent. By stratifying based on social and/or functional risk factors, the TEP noted that providers may be further incentivized to do well in treating patients with this factor, which also increases transparency for providers who provide better care for patients with this factor. The TEP added that measure developers should demonstrate appropriate use of both risk adjustment and stratification, including providing rationale and strong evidence in cases in which the measure is not risk-adjusted or stratified.

Risk Model Calibration Expectations

For this next topic, Dr. Pickering stated that the minimum standard #6 in the Technical Guidance is related to calibration. He summarized that the guidance states that risk model calibration statistics inform whether the risk adjustment model-predicted probabilities are, on average, close to the average observed probabilities. To adequately assess the impact of social and/or functional risk, risk adjustment model calibration must be examined within at-risk subpopulations. Dr. Pickering proceeded to summarize the stakeholder feedback on this topic, which requested additional guidance from the TEP as to which at-risk subpopulations by which the risk model should be calibrated.

Dr. Maddox led the discussion by asking the TEP if there were any best practices for calibration and if it follows the same stratification approaches that were just agreed upon. The TEP agreed that a similar minimum set of variables, as those with stratification, can be used to test for subgroup calibration analysis. These include disability status, ethnicity, race, and rurality. Beyond this minimum set of variables, the TEP recommended that developers assess calibration of any social and/or functional subgroups identified in the conceptual model that is specific to the measure. A TEP member commented that whether this minimum set of variables is feasible depends on data availability and distribution of each variable within subgroups.

Dr. Maddox asked if a model does not calibrate for some groups would it invalidate the full measure. Adding to that question, Dr. Pickering asked the TEP if the model is poorly calibrated for a certain subpopulation would that justify the exclusion of that risk factor from the measure. A TEP member stated that if the developer decided against the inclusion of the factor, then they would need to show that their model calibrates and has a reasonable fit within those subpopulations in which the factor exists. The TEP further clarified that the measure developer would need to determine why the model did not calibrate well and if the model were to be used, then the developer should provide a disclaimer to note the poor subpopulation results.

Dr. Pickering asked if there were any literature that can be referenced in this section of the guidance related to the decision making for calibration. A TEP member shared that there is no set literature that's

related to the specific decision making but iterated that developers should note in their data model, which subpopulation was poorly calibrated. Additionally, the TEP noted that graphical approaches may be preferred, such as plots of observed-to-expected outcomes across a broad range of expected values. However, calibration assessments of graphical approaches require methodological judgement to determine final adequacy, since calibration assessments may depend on the measure, data availability, and sample sizes for subgroups. Measure developers should examine model calibration in the populations represented by the minimum set of subgroups discussed above including disability status, ethnicity, race, and rurality. Dr. Maddox turned the discussion back to Dr. Pickering to summarize the next topic.

Key Areas of Consensus Outlined in the Executive Summary

For this topic, Dr. Pickering summarized the Executive Summary of the guidance, which draws attention to the key areas of consensus. He stated that after the conclusion of this work, this guidance advances consensus and further delineates best practices for social and functional risk adjustment within quality measurement on several key fronts:

- Developers should develop a conceptual model that illustrates the pathways between the social and/or functional risk factors, patient clinical factors, healthcare processes (e.g., care delivery behaviors, like whether certain tests were conducted, care coordination), and the outcome of interest. The rationale for risk adjustment variables must derive from the specific relationships illustrated by the conceptual model;
- Developers should 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;
- Measures adjusted for one or more social or functional risk factors should assess a stratification approach, as well as being risk-adjusted;
- Race is qualitatively different as a risk factor from other social risk factors, and risk adjustment models should generally not adjust for it;
- This guidance identifies appropriate data sources for social and functional risk adjustment; and
- This guidance provides approaches to conducting empirical analyses of risk variables as well as analyses for testing the adequacy of risk adjustment models.

Dr. Pickering asked the TEP if there were any additional considerations for these aspects of the report and turned it over to Dr. Maddox to facilitate the discussion.

The TEP largely agreed with areas of consensus, except for how race was presented. A TEP member commented that to determine racial inequities, measures should be stratified by race to review the performance across racial subgroups. Another TEP member suggested adding the framing around transparency and that risk adjustment alone may not be sufficient for providing transparency with respect to disparities and inequities. Another TEP member commented that per the [National Academies of Sciences, Engineering and Medicine \(NASEM\) report](#), stratification provides transparency, and together with risk adjustment, can address unintended incentives of avoiding at-risk patients. It was therefore suggested by TEP members to rephrase the language in the guidance to clearly point out that the primary goal of stratification is to identify and address the existing disparities.

Another TEP member disagreed with the current phrasing of the use of race, namely that "risk adjustment models should generally not adjust for it." The TEP member expressed concerns about not including race in risk adjustment models, as this may result in an increased likelihood of risk model miscalibration. As a potential solution, the same TEP member recommended setting guidelines on the

rationale for race and/or ethnicity in a model. Other TEP members agreed with the concerns about excluding race altogether, noting that a provider is responsible for not exposing a patient to racism. However, patients are exposed to racism prior to being evaluated by a health professional, which adversely impacts their health. A TEP member stated that using race as a proxy for prior exposure to racism can prevent providers from avoiding patients with unmeasured health disadvantages. Dr. Drye clarified that the language within the guidance discusses the risk adjustment models used in quality and not clinical risk prediction models.

Dr. Pickering noted that the updates to the guidance with respect to how race should be considered for risk adjustment was based on the input gained from this TEP after the presentation of the stakeholder feedback about race during web meeting #2. Dr. Pickering summarized that stakeholders expressed that race should not be used as a risk adjustment variable, as it is not clear what the variable truly represents. Additionally, stakeholders commented that race should not be used as a proxy for unmeasured social risk, as this could perpetuate the misconception that social needs and social risks are connected to race. Therefore, due to this degree of disagreement with respect to race, Dr. Pickering shared that Technical Guidance will be further refined to reflect both sides of the issue. He stated that the NQF staff will seek to update the Technical Guidance based on the discussions held today and will share this updated text with the TEP prior to the public comment period. Dr. Pickering thanked the TEP for the insightful discussion.

Public Comment

Dr. Pickering opened the web meeting for public comment. John Shaw, from Next Wave, provided a comment that agreed with the TEP's decision making for stratification and calibration, stating that the plans for incorporating health equity must include all of the must-have items (e.g., race, ethnicity, urban, rural, and disability), at a minimum. Mr. Shaw further commented, noting standard data available to developers would assist with the implantation of these items. Lastly, Mr. Shaw emphasized focusing on the cost of living in addition to income due to an additional bias in the measures related to disparities.

Sylvia Trujillo, from the Oregon Community Health Information Network (OCHIN), provided a comment by thanking and appreciating the TEP for the seriousness in the discussion of topics. Ms. Trujillo further stated that structural inequalities are a result of implicit bias and discriminatory practices. Lastly, Ms. Trujillo emphasized the Technical Guidance suggested historic societal discrimination and contemporary unintentional system bias do not address individual and current systemic racism that impact minorities.

Next Steps

Simone Bernateau, NQF analyst, began by noting TEP feedback will be incorporated into the Technical Guidance. Ms. Bernateau noted the Technical Guidance will go out for public comment from August 23 to September 14. Ms. Bernateau stated that web meeting #4 will be on October 24, 2022, in which the TEP will review and adjudicate the comments received during the public comment period. Lastly, the TEP was informed that the Technical Guidance will be finalized following web meeting #4 and will be posted on December 21, 2022.

Dr. Pickering commented that since the TEP came to an impasse on the key areas of consensus, then NQF staff will seek to update the guidance based on the discussion today and share the updated text with the TEP for review prior to public comment. This may mean convening the TEP again for another web meeting. Dr. Pickering then thanked the TEP, including the leadership of its co-chair, Dr. Maddox, the Federal Liaisons, and the members of the public for their time and participation. He then adjourned the call.