

Final Expert Panel Responses to Comments

Following are the Expert Panel's responses to the major themes identified in the comments received during the public comment period.

1. Masking disparities, masking quality problems, different standards

Commenters agreed with the recommendation that stratification was the appropriate method to identify disparities. However, some commenters objected to sociodemographic adjustment for purposes of public reporting and pay-for-performance and urged continuation of NQF's existing criteria and guidance. They expressed concerns that adjusting for sociodemographic factors masks disparities in outcomes, masks quality problems, creates different standards, and reduces the incentive to improve and reduce disparities. Other commenters noted that the analyses that are needed to include adjustment for sociodemographic factors would highlight where there are disparities (i.e., significant coefficient in a risk model). Some commenters suggested that both SDS-adjusted and stratified data be publicly reported.

Response

The term "masking disparities" is a misnomer because disparities are not visible using current clinically-adjusted measures. Masking disparities in outcomes (or processes), masking disparities in quality, and setting different standards, while related, represent distinct concerns. The Expert Panel provides two responses — one methodological and one to provide for greater transparency about disparities. Both of these are discussed in detail in the final report.

- The Expert Panel developed an in-depth discussion of the methodological basis for SDS adjustment, which is provided in section 4.
- The Expert Panel recommended that if a measure was SDS adjusted, then specifications also include instructions for stratification.

2. Evidence of harm

Some of the objections to sociodemographic adjustment were based on the perception that the primary reason for the recommendations was potential harms to disadvantaged patients related to not adjusting for sociodemographic factors and that there was insufficient evidence of such harms. Therefore, they concluded that a change in the criteria related to adjusting for sociodemographic factors is not warranted.

Response

- a. Whether to adjust for sociodemographic factors or not, and how, is first and foremost based on sound methods for quality measurement. That is, the Panel first asked the question: "Will consideration of sociodemographic adjustment improve comparability of performance between providers?" Sound measurement represents the central tenet of performance assessment and enables optimal decision making among patients, purchasers, and payers to make informed comparisons between providers and inferences about their relative quality. Sound measurement also improves perceptions of fairness among those being assessed. The majority of the Panel thinks that sociodemographic adjustment, under the conditions identified in the report and in the detailed

discussion of methods (see section 4) will produce performance measures that will provide more valid, meaningful, and fair comparisons among plans and providers on key dimensions of quality of care. This focus on best possible comparative measurement of quality is consistent with NQF's focus on quality measurement per se, rather than on the actual consequences of uses of measures (which it does not directly control).

- b. The primary evidence that is relevant to the question of whether or not to adjust for sociodemographic factors is the substantial body of evidence that demonstrates the relationship between a variety of sociodemographic factors and a variety of health outcomes (and some processes). However, it is important to note that the recommendations do not suggest sociodemographic adjustment of all performance measures, or even all outcome performance measures. The decision on whether to include sociodemographic factors needs to be made for each individual performance measure based on the conceptual and empirical relationships that exist between the factors and the outcome or process being measured as well as working through the guidelines for selecting risk factors. Therefore, a body of evidence about the relationship between sociodemographic factors and outcomes (or processes) provides only a starting point for considering sociodemographic factors as confounders and potential risk adjustment.
- c. The potential harms from not adjusting for sociodemographic factors identified in the report are potential consequences of not following accepted and sound methods to control for confounding (see the methods discussion in section 4). The Panel reviewed a number of published studies documenting harm to safety net providers, primarily through financial penalties. Fewer studies addressed potential reputational harm to providers. No studies directly assessed harm to patients under the current policies. The Panel recognized that it is a plausible, but unproven assumption that reducing revenue to financially strapped safety net organizations could eventually result in fewer resources devoted to care for disadvantaged patients resulting in worse outcomes. A few additional references related to potential harms have been identified, but that is not the primary evidence question.
- d. The Panel notes that the current policy prohibiting sociodemographic adjustment was not based on empirical evidence of benefit or harm to patients. It also notes that the National Healthcare Disparities Report produced by AHRQ shows little consistent progress in reducing healthcare disparities during the time of the current policy of prohibiting adjustment for sociodemographic factors. There also is not a body of evidence on potential harms to patients related to allowing sociodemographic adjustment (e.g., setting different standards and reducing incentives to improve).
- e. Therefore, the recommendations are based on sound principles of measurement science, and the decision of whether to adjust for sociodemographic factors needs to be made for each individual measure based on the conditions laid out in the recommendations.
- f. We have corrected the cited reference (#24 in the draft report) – the text was correct, but the correct citation: is Joynt, KE, Jha, AK (2013). Characteristics of hospitals receiving penalties under the Hospital Readmissions Reduction Program. *JAMA*, 309(4), 342-343).

3. Definition of quality, healthcare responsibility, reduce incentive to improve, impede progress on outcomes such as readmission

Some commenters thought that the discussion about what healthcare plans or providers can control or influence reflected a narrow view of healthcare quality and provider responsibility to adjust care based on sociodemographic factors. Some expressed concern that sociodemographic adjustment would impede progress that is being made on hospital readmissions and that hospitals would abandon efforts

to reduce readmissions (or potentially other important outcomes) as a result of sociodemographic adjustment.

Response

- a. The Expert Panel agrees that healthcare should be based on the characteristics of the patients served; should not lower goals or standards when providing care to disadvantaged patients; and the need to identify and reduce disparities.
- b. That said, the vast majority of comments received during the public comment period made some mention of factors outside of providers' or health plans' control that influence measured outcomes. Most outcomes are clearly a function not only of what plans and providers do, but of other factors operating at the individual, household, community, and broad societal levels. There is no widely-accepted definition of quality of care that holds doctors, hospitals, health plans, and other sorts of "providers" responsible for ALL factors leading to many measured outcomes.
- c. Sociodemographic risk adjustment does not contradict broad definitions of healthcare quality reflected in the [IOM definition of healthcare quality](#); or others such as AHRQ's: "Doing the right thing for the right patient, at the right time, in the right way to achieve the best possible results"; or CMS definition from its QI Roadmap: "Right care for every person every time." All of these definitions focus on what healthcare entities do, not about what society does or does not do.
- d. Risk adjustment for certain factors does not absolve providers/plans from the responsibility to use interventions appropriate for those factors when present in the patients served whether clinical factors (e.g., recognizing and addressing comorbidities) or sociodemographic factors (e.g., recognizing and addressing non-English speaking persons). This holds whether clinical factors or sociodemographic related factors are being considered for adjustment.
- e. Adjustment for sociodemographic factors when indicated improves comparability among providers/plans. It does not place a limit on the scope of interventions that could be used to mitigate the effects of sociodemographic factors such as the number of language translations or interpreters available or "discharge clinics" for patients without primary care providers. Risk adjustment creates a "level playing field" so that differences across providers/plans in addressing or not addressing the sociodemographic factors will be reflected in the adjusted performance measure scores.
- f. Risk adjustment does change the estimate of the provider's performance (either up or down) depending on the proportion of patients in the SDS categories. This is appropriate in the context of the question: ***how would the outcomes of various units compare if hypothetically they had the same mix of patients?*** (See section 4.) However, if the question is: ***how do the outcomes of patients with different characteristics compare (either within an individual unit or at the population level)?*** then a different analysis is indicated. As recommended by the Expert Panel and in prior NQF projects, identifying disparities in either outcomes or processes requires additional information and analysis (e.g., stratification by relevant sociodemographic characteristic).
- g. Adjustment for sociodemographic factors when indicated does not necessarily remove the focus of improvement or the need to work collaboratively with other settings, depending on the performance measure. By measuring and comparing performance on risk-adjusted rates, providers/plans, and others can identify when performance is lagging and providers/plans that are achieving excellent performance. For improvement, providers/plans always need to examine their own data stratified by relevant clinical and/or sociodemographic characteristics to identify patients who are and are not achieving desired outcomes and potential strategies to improve. Additionally, risk adjustment procedures should be updated on a periodic basis so that improvements are reflected in updated model coefficients.

4. Methods

Some comments were about methods or description of methods in the report:

- ***“Not primarily mediated by quality” should not be a requirement for selecting risk factors***
Comments by a statistician and an epidemiologist caution against focusing on causal pathways. The statisticians on the Panel also recommended that this language is not needed. It is difficult to define in order to operationalize and therefore, could potentially add burden to the measure development process.

Response

- a. Based on epidemiologic principles related to confounding and statistical theory of causal inference, the language “not primarily mediated by quality” is not needed (see methods discussion in section 4) and has been omitted from the revised recommendations. The decision on whether to include sociodemographic factors needs to be made for each individual performance measure based on the conceptual and empirical relationships that exist between the factors and the outcome as well as working through the guidelines for selecting risk factors. However, an assessment of a conceptual relationship between an SDS factor and outcome of interest includes a consideration of whether the effect of the SDS is primarily mediated by the quality of care delivered. This is discussed in section 6.

- ***Disagree with characterization of sociodemographic adjustment making more “accurate” or “correct” conclusions and suggest language that risk adjustment improves comparability***

Response

- b. One of the core principles used the language “avoid making incorrect inferences about performance” and is an appropriate statement related to risk adjustment. Making correct conclusions is a logical statement of the same concept from a positive perspective. Language used to describe validity (to which risk adjustment relates) often refers to “accurate” and “correct” but varies by disciplines and preference. The term “accurate” is also used sometimes to indicate precision and could be confusing. The references to “accurate” have been replaced with the terms: avoid incorrect inferences, improve comparability, and make unbiased estimates (statistical term used in the methods report) depending on the context.

5. Implementation is the issue, not measurement

Some of the objections to the recommendations were based on the perspective that the issue (harm to providers or patients through lack of adjustment) was really about how the measures were used in pay-for-performance programs and not about measurement per se. Some suggested alternative ways to structure incentive programs. Some advocated for peer group comparisons as recommended by MedPAC for the hospital readmission measure. However, some other commenters suggested that the alternative of peer groups for comparison explicitly accepts or creates different standards for plans or providers grouped by a sociodemographic variable. One commenter noted specific mechanisms for adjusting payment for services based on higher needs related to sociodemographic factors and therefore, adjustment for performance measures could result in overpayment.

Response

- a. The panel focused primarily on the question of whether consideration of sociodemographic adjustment improves the performance measure for comparisons and avoids incorrect inferences about quality. Although concerns about the impact of payment incentive programs might have been the impetus to re-examine NQF's policy on adjusting for sociodemographic factors, the primary basis for the recommendations is that they are consistent with accepted practices and guidelines for selecting risk factors for performance measurement and epidemiologic and statistical approaches to handle confounding in order to enable comparisons and avoid incorrect inferences about quality regardless of the specific accountability application.
- b. The concerns of the Panel have not just been limited to issues of payment incentive programs. Rather, the concerns of the Panel are also set in the context of public reporting and the validity of inferences or comparisons made with performance measures that are not adjusted for sociodemographic factors when appropriate. Alternatives to adjustment that may be useful in pay-for-performance contexts do not address a deeper concern that failing to consider sociodemographic adjustment can yield performance measures that may be fundamentally misleading to patients, consumers, purchasers, payers, and regulators who are engaged in making comparisons among plans or providers.
- c. Appropriate adjustment for sociodemographic factors may not be sufficient to address the financial issues of safety net providers/plans; however, the performance measures used in such programs should provide an unbiased estimate (i.e., without systematic deviation from the true value) of performance on the quality measure for the entity being measured and compared.
- d. Peer groups for comparison: As noted by some commenters, unlike model-based adjustment, this approach does have the potential to mask quality differences. One commenter elaborated: "The two approaches are fundamentally different in that risk adjustment adjusts for the distribution of patient characteristics (such as poverty), while peer group comparison adjusts for unit characteristics. For example, if comparisons are made within a peer group of hospitals that have trouble providing high quality care because they are under resourced and poorly reimbursed, we might say a hospital is superior to its peer group even though the same patients would have received superior care at another hospital outside the group. Conversely if a hospital is superior in risk adjusted scores, it suggests that the same group of patients would do better there than at another hospital. Peer group comparison may have a place as a tool of the incentive system rather than as part of the construction of the measure itself."
- e. Stratification: The Expert Panel discussed the statistical limitations, mainly in the form of small sample sizes for computing performance scores for each stratum for an individual physician, or small physician group, or small hospital. One of the commenters elaborated: "Sample sizes for some measures adequate for estimation (with adequate reliability) of a single measure for a unit, but not for separate estimation of measures for strata (subgroups), especially when some strata have only sparse representation in some units. This is unlikely to be a problem, however, for model-based statistical adjustment, since model parameters may be estimated from the combined data from a multitude of units. Furthermore, these model parameters give a summary measure of within-unit disparities that typically is more sensitive than what can be discerned from perusing a set of stratified results."

6. Burden to developers, guidance to developers

Some of the objections were based on burden to measure developers and concern that developers would not develop performance measures that required sociodemographic risk adjustment. Other commenters cautioned about potential developer burden and suggested more guidance for developers would be needed.

Response

- a. Risk adjustment is a complex and nuanced area of methodology and requires expertise that may not be present in all measure developers. It is difficult to anticipate all possible scenarios to create more prescriptive directives and rules. Measure developers need the flexibility to use the methods that are indicated in a particular situation.
- b. Although plans and providers may not directly pay measure developers for their work, the support for, and potentially greater acceptance of, sociodemographically adjusted measures by plans and providers will give some measure developers an opportunity rather than a burden.
- c. The measure submission questions should guide what measure developers are expected to present for review and evaluation.
- d. Nothing in the Panel's recommendations asks or demands that measure developers collect or analyze primary data. The obligation on measure developers is presumably exactly like the obligation that already exists with regard to clinical variables used for adjustment. Measure developers will be obliged to recognize, and incorporate when possible, existing valid empirical data on the association between sociodemographic factors and "outcomes" (or some processes). When such data exist, developers may have to do more work than under the current policy prohibiting sociodemographic risk adjustment; this additional work should not be a barrier to the creation of measures that, in use, will provide more valid and informative comparisons among plans and providers.
- e. Initially, data limitations may constrain what is feasible, and NQF Committees will need to recognize that. If the recommendations are implemented, SDS data will improve over time.

7. Data burden, feasibility

Some commenters saw sociodemographic data limitations as a reason to delay implementation. Other commenters cautioned about the potential of making data collection too burdensome. Some commenters noted that potential adjustment for sociodemographic factors would provide incentive to collect the necessary data. Some commenters noted other efforts related to data on sociodemographic factors, specifically recent IOM work [Capturing Social and Behavioral Domains in Electronic Health Records: Phase 1](#).

Response

- a. Initially, data limitations may constrain what is feasible either in the sense of development and testing of adjustment models or in the sense of using an SDS-adjusted measure in public reporting or pay-for-performance. NQF Committees will need to recognize that. The collection and availability of sociodemographic data are likely to advance as follows:
 - Initially, developers will primarily need to use variables readily available in existing data sets (e.g., Medicaid status); then
 - patient or member address for geocoding to census tract data; then

- standard definitions and data collection processes as defined and supported by groups such as AHRQ, IOM, and CMS.

8. Additional sociodemographic factors, community factors

Some commenters suggested additional factors that should be considered or that more attention should have been given to community-level factors.

Response

- Potential sociodemographic factors were identified in the report, but currently there is no basis for being more prescriptive about specific risk factors, especially when decisions about risk adjustment need to be determined for each individual measure.
- Sociodemographic factors could be obtained from three sources:
 - sociodemographic data collected from each individual (e.g., race/ethnicity, literacy, homelessness, English proficiency, marital status, etc.);
 - census variables obtained through address geocoding usually at the census tract level, but could be identified for other levels like ZIP Code (e.g., percent below poverty level, percent employed, average education level); and
 - community resource variables that come from sources other than census data (e.g., strength of primary care network in a community, availability of visiting home nurses, Meals on Wheels, public transportation, community health centers, etc.).
- The Panel agrees that community factors such as availability of public transportation, size and strength of community health center network, availability of primary care, availability of support services like Meals on Wheels, etc. can have a profound effect on patient outcomes. Risk adjustment addresses patient characteristics (see section 4) so community characteristics need to be specific to each patient. Community characteristics could be assigned to each individual patient (e.g., percent poverty or public transportation in community **where the patient resides**). Generally, characteristics of the healthcare unit are not considered patient risk factors (e.g., percent poverty or availability of public transportation in the community **where the unit is located or for the patient population served**). However, some community factors such as public funding of safety net providers have implications for the capacity of healthcare units to deliver quality service and policy implications for the response to performance assessment.

9. Implementing the recommendations and monitoring impact

Some commenters suggested more research, incremental approaches to implementation, and monitoring impact. Other commenters suggested immediate implementation and review of endorsed measures to identify those that might require an ad hoc review.

- Adoption of the Panel’s recommendations about sociodemographic adjustment and stratification will inevitably be “incremental.” That is, measures currently in use will not have to be considered for sociodemographic adjustment until the next review cycles for those measures come up. Some measures for which a strong conceptual argument for adjustment exists will not be able to be implemented with sociodemographic adjustment because data constraints prevent development and validation of an adjustment model. For other measures, the data may be available to develop and validate a model, but not be available to routine use in a large population of plans or providers.
- The limited evidence available to date about the effects of sociodemographic adjustment suggests that the effects will not be profound. That is, providers or plans may move to some extent up or

down in relative rankings, but “good” will not instantly become “bad” and vice-versa. Effects of adjustment will likely be modest, based on analyses that have been done and reported to date, but could be substantial for some healthcare units.

- c. To address these concerns, the Expert Panel recommended a transition period to monitor implementation (Recommendation #2) and that NQF appoint a new standing committee focused on disparities to review implementation and monitor o unintended consequences (Recommendation #3).

10. Clarifications

Some comments requested specific clarifications or indicated the need for clarification. Following are some specific clarifications.

Are health plans included?

Are cost and resource use measures included?

Some comments seemed to imply that all performance measures would be adjusted for sociodemographic factors.

Response

- a. The recommendations apply to performance measurement for any setting or unit of analysis, including health plans.
- b. The recommendations apply to outcome performance measures (including cost and resource use and PRO-based performance measures) and some process measures depending on the specific circumstances. The recommendations are purposely not prescriptive in terms of factors and methods — that needs to be determined for each individual measure.
- c. The recommendations do not mean that all performance measures should be adjusted for sociodemographic factors — that has to be determined for each individual performance measure. The Panel’s recommendations and supporting text are clear that the recommendation about sociodemographic adjustment applies only in specific circumstances. Examples of measures that would generally not be adjusted are provided in the report.

11. Opposed to NQF having a role in guidance on implementation of endorsed performance measures (Recommendation 7)

Five commenters who were in support of most of the recommendations did not agree that NQF should have a role in providing guidance on implementation and use of endorsed performance measures. The commenters think that this is outside NQF’s role for endorsing performance measures and overlaps with the role of the Measure Applications Partnership (MAP).

Response

How a measure is implemented involves multiple decisions that could affect the validity of conclusions (inferences) made about quality of care and potential unintended consequences. The recommendation is for NQF to consider expanding its role to include guidance on implementation of performance measures. This will require NQF’s decisionmaking bodies (CSAC and Board) to explore the pros and cons and implications for endorsement and measure selection for specific program uses. This fits with work

already underway at NQF to explore ways to make the measure endorsement and measure selection processes more coherent and efficient.