NATIONAL QUALITY FORUM

Risk Adjustment and Sociodemographic Factors
Draft Recommendations and Rationale

WORKING DRAFT 01/31/14

NOTE: The purpose of this document is for the Expert Panel to review draft recommendations and rationales that emerged from the in-person meeting. It is not intended as a stand-alone report and additional descriptions and explanations will be included in the full report. This will be the subject of our follow-up call on 2/10.

Core Principles

1. Outcome performance measurement is critical to the aims of the national quality strategy.
2. Outcomes may be influenced by patient health status, clinical, and sociodemographic factors, in addition to the quality and effectiveness of healthcare services, treatments, and interventions.
3. When used in accountability applications,[1] performance measures that are influenced by factors other than the care received, particularly outcomes, need to be adjusted for relevant differences in case mix to avoid incorrect inferences about performance.
4. Disparities in health and healthcare[2] should be identified and reduced.
5. Performance measurement should not lead to increased disparities in health and healthcare. [2]
6. Risk adjustment may be constrained by data limitations and data collection burden.

Risk Adjustment – The process of controlling or accounting for patient-related factors when examining outcomes of care, regardless of context.[4] Generally, process performance measures are not risk-adjusted; however, risk adjustment of process performance measures could be considered when the process in question is at least partially under the control of the patient, or dependent on factors in the patient, family, or community that the provider can't influence.

Adjustment methods include:

- Comparison of observed to expected outcomes for an accountable entity
    - Indirect standardization where the expected number of outcomes are determined by applying stratum-specific rates determined from all patients to the number of cases in each stratum for each provider
    - Extension to multivariable statistical models

[1] **Accountability Applications** – Use of performance results about identifiable, accountable entities to make judgments and decisions as a consequence of performance, such as reward, recognition, punishment, payment, or selection (e.g., public reporting, accreditation, licensure, professional certification, health information technology incentives, performance-based payment, network inclusion/exclusion).

[2] **Health Disparity** – Healthy People 2020 defines a health disparity as "a particular type of health difference that is closely linked with social, economic, and/or environmental disadvantage. Health disparities adversely affect groups of people who have systematically experienced greater obstacles to health based on their racial or ethnic group; religion; socioeconomic status; gender; age; mental health; cognitive, sensory, or physical disability; sexual orientation or gender identity; geographic location; or other characteristics historically linked to discrimination or exclusion."

Healthcare disparity – Differences in health care quality, access, and outcomes adversely affecting members of racial and ethnic minority groups and socially disadvantaged populations.

- Stratification of outcome results for patients in different risk categories (e.g., severity of illness, household income) within each accountable entity. Each provider has a performance score for each category and comparisons are made across categories rather than an overall performance score.
- Organizational stratification to create peer groups or organizations with a similar patient mix (e.g., organizations ranked in quintiles or deciles based on proportion of patients below 138% of poverty level). This stratification does not affect the calculation of the performance score and occurs during implementation (e.g., public reporting, pay-for-performance).
- Combination of statistical risk model and stratification (e.g., statistical adjustment for clinical factors and stratification for sociodemographic factors; different statistical models for each stratum that are used to compute one overall performance score).

_____

Recommendation 1:
Two versions of the first recommendation were discussed.

Recommendation 1-Option A: NQF should require that sociodemographic factors be considered for adjusting outcome performance measures and require information about what factors, how they were considered, analyses, and rationale for adjusting or not adjusting (addressed in the following recommendations).	**Recommendation 1-Option B:** NQF should require that outcome performance measures be adjusted for sociodemographic factors unless justification is presented for not adjusting for sociodemographic factors.
All the Panel members supported this language.	**About half the Panel members supported more prescriptive language.**
Rationale: Sociodemographic factors are appropriate for consideration for adjustment of outcome performance measures. This requirement signals the importance of considering sociodemographic factors for potential adjustment. The requirement to submit information for review in the endorsement process ensures an open and transparent process for the endorsement decision. This requirement is more explicit than for clinical or health status factors in order to raise awareness.	**Rationale:** Sociodemographic factors are not only appropriate for consideration, but almost always necessary for adjustment of outcome performance measures. The default is to include sociodemographic factors unless justification is presented for not including them. This requirement is more prescriptive than Option A.

_____

Recommendation 2: The same considerations for selecting clinical and health status risk factors for adjustment of performance measures should be applied to sociodemographic factors.

Rationale: The considerations for selecting clinical risk factors apply equally well to sociodemographic factors. See Table 1. The purpose of Table 1 is to examine whether there are any differences when considering sociodemographic risk factors. Not all risk factors will meet every consideration. Selecting risk factors and developing a model is an iterative process.

NATIONAL QUALITY FORUM

Table 1. Considerations for Selecting Risk Factors for Outcome Performance Measures

Consideration	Rationale	Clinical/Health Status Factors	Sociodemographic Factors
		e.g., comorbidity; severity of illness; patient-reported health status	e.g., income; education; language proficiency
Clinical/conceptual relationship with the outcome of interest including the accountable entity	Begin with conceptual model informed by research and experience	✓	✓
Variation in prevalence of the factor across the measured entities	If no variation, it will affect all entities equally	✓	✓
Empirical association with the outcome of interest	To confirm conceptual relationship	✓	✓
Not confounded with quality of care – risk factors should be:	Trying to isolate differences in quality of care	✓	✓
• Present at the start of care and	Ensures not a result of care provided	✓	✓
• Should not represent the quality of care provided (e.g., treatments, expertise of staff)	Although could explain variation in outcome, trying to isolate differences due to the care provided	✓	✓
Resistant to manipulation or gaming – generally, a diagnosis or assessment data (e.g., functional status score) is considered less susceptible to manipulation than a clinical procedure or treatment (e.g., physical therapy).	Ensures validity of performance score as representing quality of care (vs. for example, upcoding)	✓	✓
Accurate data that can be reliably and feasibly captured	Data limitations often represent a practical constraint to what factors are included in risk models	✓	✓
Contribution of unique variation in the outcome (i.e., not redundant or highly correlated with another risk factor)	Prevent overfitting and unstable estimates, or coefficients that appear to be in the wrong direction; reduce data collection burden *If redundant, which one is retained?*	✓	✓
Potentially, improvement in risk model metrics of discrimination and/or calibration and sustained with cross-validation	Change in R-squared or C-statistic may not be significant, but calibration at different deciles of risk might improve (*could you analyze calibration for different*	✓	✓

Consideration	Rationale	Clinical/Health Status Factors	Sociodemographic Factors
	deciles based on some sociodemographic variable?) May not appear to be a big change but could represent meaningful differences in terms of the outcome (e.g., lives, dollars) Order of entry into a model may influence this result		
Potentially, face validity and acceptability	Some factors may not be indicated empirically, but could improve acceptability – need to weigh against negative impact on model, feasibility and burden of data collection	✓	✓

Recommendation 3: When submitting outcome performance measures for potential NQF endorsement, the following information about consideration of sociodemographic factors should be provided:

- Conceptual description - causal pathway between factors and outcome
  - Informed by review of literature, content experts
- Sociodemographic variables that were considered including different level of variables and approaches
  - Patient-level
    - Patient-reported (e.g., income, education, language)
    - Proxy (e.g., based on patient address, use census tract data to assign to a category of income, education, etc.)
  - Provider-level
    - Aggregation of patient-level data (e.g., proportion of patients at 200% or less of federal poverty level)
    - Community variables for service area (e.g., percent vacant housing)
- Analyses and interpretation resulting in decision to include or not
  - Prevalence of the factor across measured entities
  - Empirical association with the outcome
  - Contribution of unique variation in the outcome
  - Effect on performance score (including effect on ranking) with and without adjustment for sociodemographic factors

Rationale: The developer's decisions regarding sociodemographic factors should be transparent and open to review and evaluation.

Recommendation 4: The methods for adjustment, analyses, interpretations, and decisions should be justified, demonstrated as meeting criteria, and submitted for evaluation.

Rationale: NQF should not be prescriptive regarding methods for adjustment because there are multiple approaches to achieve the same goals and being too prescriptive can hinder progress in measurement approaches. However, if there are particular methods or approaches that should not be used, those should be identified. Steering committees and stakeholders need to have all information needed to evaluate performance measures for endorsement.

Recommendation 5: Sociodemographic factors that should always be considered for adjusting outcome performance measures include: income, education, homelessness, English language proficiency, and insurance status. Selection of specific variables will depend on available data. (See Table 2)

- Consider different levels of variables and approaches as described in Recommendation 4.
- Consider the use of census tract data matched to patient address, especially for data that are difficult to collect on an individual basis (e.g., income – difficult to obtain in a healthcare encounter and household income is likely to be a better indicator)
- For national performance measures, variables used for sociodemographic adjustment factors should have standard definitions (e.g., Medicaid status may not provide an accurate reflection of income due to variations in eligibility by state)
- Race and ethnicity should not be used as proxies for socioeconomic status

Rationale: Currently, there is no agreed-upon and standardized set of sociodemographic variables that can be required for use in performance measurement. This will be an evolving area and will need to be revisited as more experience is gained and best practices identified.

Table 2. Sociodemographic Factors

Factors/concepts (specific variables)	**PROs**	**CONs**	**Caveats**
Factors that always should be considered, depending on: data availability and the outcome being measured			
Income	• Allows for use of various ranges	• Hard to collect privately (e.g., in clinician office) • Not easily collected with a single question • May not be an acceptable question to patients • Meaning is not geographically consistent	
Federal poverty level	• Definition is standard • Being used under ACA • Researchers are used to	• Doesn't include receipt of other benefits (e.g., food stamps)	

Factors/concepts (specific variables)	PROs	CONs	Caveats
	using it		
Household income	• May be more meaningful than individual income	• May need to also capture household size	
Medicaid status as proxy	• Relatively easy to collect in claims data	• Eligibility not consistent across states	• Potentially becomes more useful as more States expand Medicaid to 138% FPL
Social Security Supplemental Income (SSI)		• Correlated with Medicaid status, but not consistently across states	• In many states, receipt of SSI automatically makes eligible for Medicaid
Education	• Perceived to be valid (i.e., less socially desirable response than for income) • Definitions fairly consistent across various subgroups (e.g., answers from immigrants comparable to those from others) • Fairly stable across time, at least after a certain age • Moderate correlations with other SES measures, e.g. occupation and income	• Possibly not widely collected • If collected (e.g., in EHR) may not be easily retrievable	
Homelessness	• Strongly associated with health outcomes • Measures something "beyond" income	• Multiple definitions • Data often not collected • Status can change	• Prevalence tends to cluster among safety net providers
Housing instability	• May be better indicator than homelessness which can change	• More difficult to define than homelessness	
English proficiency	• Standard definition does exist • Tied to need for translation services/other resource need and therefore should be collected • Increasingly being collected		
Insurance status			

Factors/concepts (specific variables)	PROs	CONs	Caveats
Medicaid status		• Not consistent across states	
No insurance			• Difficult to capture information about these patients (particularly if using claims data)

Neighborhood-level data could be used as proxy for individual data or as an aggregate-level contextual variable	• Many variables available from Census data
	  • Income
	  • Education
	  • Immigration status
		• Census data do not include all potentially important variables
		• Residential heterogeneity will affect whether it's a good proxy.
		• Heterogeneity may differ based on levels of socioeconomic segregation and potentially population density.
		• Requires geocoding for Census Tract and smaller areas.	

Contextual - Proportion vacant housing			
Contextual - Grocery stores			
Contextual- Crime rate			
Other factors that may be considered			
Factors/concepts (specific variables)	**PROs**	**CONs**	**Caveats**
Social support	• Some brief items have been used in previous research • Captures something that other variables do not	• Disagreement about how to measure • Not consistently measured	
Living alone	• Available in OASIS data for home health	• Directionality may not be consistent if it is primarily an indicator of good health and function	• People who are healthy but living alone still may struggle with post-surgical recovery
Marital status	• Often collected		
Occupation	• May capture other concepts (e.g., environmental exposures)	• Multiple definitions • Potentially large data collection burden due to the complexity of the concept • Marginal value (i.e., over and	

Factors/concepts (specific variables)	PROs	CONs	Caveats
		above that contributed through use of other variables) may be limited • Unclear how to handle certain population subgroups (e.g., retirees, students, homemakers)	
Employment status		• Simple yes/no does not reflect desire/happiness with situation (e.g., retires may be happy to be unemployed)	
Literacy	• This concept may also be able to capture health literacy	• No standardized definitions • May be easy to game • Correlated with education	If the correlation with education was high, then education could be used.
Health literacy	• Potentially more relevant to health care • Three-item and single-item validated questions exist	• Not consistently collected/ available	•
Local/state funding availability (e.g., tax base)	• Affect resources available to safety net providers beyond insurance	• Data not easily collected/ available	• Risk for unintended consequences (setting a lower standard for poorly supported institutions might send the wrong messages to tax payers)
Factors that should not be considered for proxies of SES			
Race/ ethnicity	• Correlated with SES and may be more available than other variables	• May be more correlated with bias	• Should not be used as proxy for SES

Recommendation 6: Measures of **processes** that are not primarily under the control of the provider should be considered for potential adjustment for sociodemographic factors (e.g., patient accepting vaccination, getting prescription filled vs. administering the correct antibiotic to prevent surgical site infection).

Rationale: The same issues regarding impact of sociodemographic factors on outcomes apply to some processes of care.

Question: Given, the above recommendations, does this mean that clinical variables also should be considered for adjusting process performance measures? Why or why not?

Recommendation 7: NQF should expand its role to include guidance on implementation of performance measures. Possibilities to explore include:

- guidance for each measure as part of the endorsement process or
- standards for different accountability applications.

Rationale: 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. For example, cut points based on rankings of performance scores without confidence intervals could result in different classifications (conclusions) about quality without any significant difference in performance in performance for providers above or below a cut point (e.g., confidence intervals for scores above and below a cut point may overlap). The structure of a pay-for-performance program could potentially worsen disparities if providers with more limited resources receive penalties and have fewer resources to provide care and improve. For example, use of pay for improvement minimizes may minimize concerns about risk adjustment adequacy and the potential for worsening disparities while maintaining an incentive for improvement.

Question: Was there a preference for one of these approaches? If so, why?

Recommendation 8: Developers should submit guidance on the intended use of the performance measure that is subject to review and evaluation.

Rationale: NQF already requires information on the intended use of performance measures and requires implementation for accountability and public reporting within specified time periods after endorsement. This would explicitly require a review of the intended use. The developer has detailed knowledge about the limitations of the performance measure that could impact its use in accountability applications.

Recommendation 9: NQF should clarify that endorsement of a performance measure is for a specific context as specified and tested for a specific patient population, data source, setting, and level of analysis. Use should not be expanded without review and usually additional testing.

Rationale: This is implicit in the current NQF criteria and process for endorsing a measure as specified and tested. However, it should be clearly stated that expansions to additional patient populations, data sources, settings, or levels of analyses are not endorsed and requires an ad hoc review.

_____

Please note this recommendation is followed by a more extensive discussion, which also has implications for some of the other recommendations.

Recommendation 10: When sociodemographic factors affect the outcome, adjustment for sociodemographic factors of performance scores used for accountability should preserve the ability to identify and address disparities.[2]

Examples of approaches for accomplishing this include:

- Performance scores should first be adjusted for clinical and health status factors (statistical models or indirect standardization would not include sociodemographic variables); then reported in the context of peer groups based on key sociodemographic factors
- After clinical adjustment, two scores should be reported – with and without additional sociodemographic adjustment.
- Sociodemographic adjustment should be used in pay-for-performance applications

Rationale: Generally, it is preferable to have one performance score for simplicity and to prevent confusion interpreting the meaning of different scores. However, sociodemographic adjustment should be used in pay-for-performance applications because there is the possible unintended consequence of diverting resources away from providers who are disproportionately providing services to disadvantaged populations. However, performance scores that are not adjusted for sociodemographic factors are important to highlight the differences in performance to drive improvement.

Question: This recommendation raises some of the original questions. Referring to Table 1, if there is no difference between sociodemographic factors and clinical and health status factors, what is the rationale for treating them differently in this recommendation? If a performance measure is intended to be a reliable and valid reflection of quality, why is not reliable and valid for public reporting or pay-for-performance, or accountability applications? This is further explored in the following discussion.

Discussion:

Much of the controversy about adjusting for sociodemographic factors is based on the assumption that adjusting for factors such as race and SES masks potential disparities in care related to those factors. The discussion to date and preliminary recommendations in some ways continue to accept that assumption but attempt to ameliorate the potential negative impact for accountability applications – i.e., report scores without adjustment for sociodemographic factors to identify disparities and with adjustment for sociodemographic factors for accountability applications, particularly pay-for-performance. However, without knowing the case mix of a measured entity it is not possible to make an inference about disparities. That is, if you did not know that a provider served a low-income population, a high readmission rate would not lead to a conclusion it was indicative of disparities in health or healthcare for low-income patients. Therefore, some of the recommendations and rationales may seem inconsistent. If the considerations for clinical and health status factors apply equally to sociodemographic factors (see recommendation 2 and table 1), then why should they be treated differently as indicated in recommendation 8?

The following discussion is intended to stimulate a critical analysis of the assumption in order to assist with final recommendations. First some questions are posed to examine the logic, and then a numeric illustration is presented.

Question Set 1: If the interpretation of the expected outcome computed during risk adjustment is "the outcome expected if the average provider's performance was applied to my mix of patients," then why not include sociodemographic factors (if they meet the considerations outlined in Table 1)? If sociodemographic factors are not included, can the model accurately reflect the average provider's performance?

Question Set 2: If including clinical and health status factors in a risk model does not mask differences in quality of care for patients of higher severity, then why would including sociodemographic factors mask differences in the quality of care for patients with higher sociodemographic disadvantage? Severity of CHF might be included as a risk factor, but it isn't considered to "mask" differences in quality; if providers do a poorer job than average in caring for patients with the highest severity, won't that be reflected in the performance score? If sociodemographic factors were included in a risk model, wouldn't providers doing a poorer job than average in caring for patients with certain sociodemographic factors also have poorer performance scores? If poorer quality is correlated with sociodemographic factors (e.g., biased care), does that make a difference?

Question Set 3: If a clinical risk factor such as severity of illness could possibly be associated with poor quality, then why can't a sociodemographic risk factor that may be associated with poor quality be included? For example, poorer outcomes for patients with higher severity may be potentially mediated by aspects of quality such as experience, skill, or available technology. Likewise, poorer outcomes for patients with higher disadvantage may be potentially mediated by aspects of quality such as not providing instructions in primary language, personal biases/assumptions about patients based with particular sociodemographic characteristics).

Illustration

The following numeric illustration (Table 3) is based on one approach to adjustment – indirect standardization as discussed in Iezzoni's book on risk adjustment. It is used to explore the assumption that indirect standardization and its extension to multivariable statistical models will mask disparities and thus hide the need for improvement. For purposes of this illustration, assume that the initial provider death rate is already adjusted for clinical and health status factors.

Provider A's case mix is more favorable (i.e., a greater proportion of low risk patients) than that of Providers B and C and its rate before sociodemographic adjustment is better (2.8%). However Provider A has poorer performance with high risk patients and its sociodemographic-adjusted rate ends up being worse (higher) than Providers B and C. Providers B and C have the same case mix in terms of high and low risk patients. However provider B's performance with high risk patients (5%) is better than Provider C's performance (7%). Provider B ends up with a better (lower) sociodemographic-adjusted rate than Provider C (3% vs. 4.18%). Thus, even though the performance score is adjusted for sociodemographic factors, it is possible to identify differences in performance in caring for high sociodemographic risk patients.

The same limitations of risk adjusting for clinical factors applies to this example for sociodemographic factors – when patient mix affects outcomes and differs widely across providers. Therefore, adjusting

the performance score does not rule out the potential need for creating peer groups in various accountability applications.

Question: Does this illustration correctly represent adjustment using indirect standardization? Will sociodemographic-adjusted scores mask disparities? If scores are not adjusted for sociodemographic factors, are disparities more evident? Unless the case mix is reported with the performance score, it does not seem any conclusion about disparities can be made. Given the information in Table 3, does Provider A's rate of 2.8% or 4.67% mask disparities in care?

Table 3. Numeric Illustration (*Note: This is one approach to adjustment. An illustration using regression model coefficients could also be developed.*)

	All Patients		**Provider A**		**Provider B**		**Provider C**	
Sociodemographic Risk Category	Pt mix %	Death%	n	Clinically adjusted Death%	n	Clinically adjusted Death%	n	Clinically adjusted Death%
Low sociodemographic risk (higher education, income, English proficiency)	50%	1.00%	800	1.00%	200	1.00%	200	1.25%
High sociodemographic risk (lower education, income, homeless, non-English)	50%	5.00%	200	10.00%	800	5.00%	800	7.00%
Performance			**n deaths**	**Death%**	**n deaths**	**Death%**	**n deaths**	**Death%**
Observed deaths and rate (O)		**3.00%**	28	**2.80%**	42	**4.20%**	58.5	**5.85%**
Expected deaths (E)			18		42		42	
Standard mortality ratio (SMR) = O/E			1.56		1.00		1.39	
Risk adjusted mortality accounting for sociodemographic risk) = rate for all pts * SMR				**4.67%**		**3.00%**		**4.18%**