Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors

TECHNICAL REPORT

July 2, 2014

NATIONAL QUALITY FORUM

Contents

Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors

EXECUTIVE SUMMARY

Introduction

There is a large body of evidence that various sociodemographic factors influence outcomes, and thus influence results on outcome performance measures. ***Sociodemographic Status (SDS)*** refers to a variety of socioeconomic (e.g., income, education, occupation) and demographic factors (e.g., age, race, ethnicity, primary language). There also is a large body of evidence that there are disparities in health and healthcare related to some sociodemographic factors. Given the evidence, the overarching question addressed in this project is, "What, if anything should be done about sociodemographic factors in relation to outcome performance measurement?"

NQF endorses performance measures that are intended for use in accountability applications such as public reporting and pay-for-performance. In this context, the overall performance measure score is used to make a conclusion about a healthcare unit's (a unit refers to an hospital, health plan, practice or other unit that is being assessed) quality in relation to other units or some other comparator such as average performance. The general question being addressed is: ***how would the performance of various units compare if hypothetically they had the same mix of patients*****?** That is, the measure scores are used to inform decisions among those seeking care regarding which units have better quality, among purchaser who pay for care, among payers regarding bonuses or penalties, or among networks for contracting, etc. Such comparisons should be affected as little as possible by factors other than quality of care, including patient characteristics.

Because healthcare outcomes are a function of patient attributes (including SDS) as well as the care received, and patients are not randomly assigned to units for healthcare services so that all have the same mix of patients, risk adjustment is essential to ensuring an "apples with apples" comparison when examining outcome performance in real-world settings. ***Risk adjustment (also known as case-mix adjustment)*** refers to statistical methods to control or account for patient-related factors when computing performance measure scores; methods include multivariable modeling, indirect standardization, or direct standardization. These methods can be used to produce a ratio of observed to expected, a risk-adjusted rate, or other estimate of performance. Risk adjusting outcome performance measures to account for differences in patient health status and clinical factors (e.g., co-morbidities, severity of illness) that are present at the start of care is widely accepted. This report explores also adjusting performance measures for sociodemographic status (SDS) when appropriate.

Core Principles

The Expert Panel on Risk Adjustment for Sociodemographic Factors agreed on a set of core principles to ground its recommendations.

1. Outcome performance measurement is critical to the aims of the National Quality Strategy.
2. Performance measurement and risk adjustment must be based on sound measurement science.
3. Disparities in health and healthcare should be identified and reduced.

4. Performance measurement should not lead to increased disparities in health and healthcare.
5. 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.
6. When used in accountability applications, performance measures that are influenced by factors other than the care received, particularly outcomes, need to be adjusted for relevant differences in patient case mix to avoid incorrect inferences about performance.
7. Risk adjustment may be constrained by data limitations and data collection burden.
8. The methods, factors, and rationale for risk adjustment should be transparent.

Recommendations

The Expert Panel made ten recommendations. The recommendations may apply to outcome performance measures (including resource use and patient-reported outcomes) and some process performance measures. However, ***each performance measure must be assessed individually to determine appropriateness of SDS adjustment***. The recommendations may apply to any level of analysis including health plans, facilities, individual clinicians, accountable care organizations, etc.

Although the recommendations to adjust for sociodemographic factors when indicated are grounded in sound measurement science methods and principles, the Expert Panel addressed concerns raised in the public comment period about appropriateness of adjusting for SDS in three substantial ways:

- requiring measure specifications for stratification to identify disparities if a performance measure is SDS-adjusted;
- recommending a transition period during which a clinically-adjusted version of the performance measure would be specified and available only for comparison purposes to the SDS-adjusted score; and
- recommending an NQF standing Disparities Committee to monitor implementation of the revised policy as well as ensure continuing attention to disparities.

Recommendations Related to NQF Criteria and Processes Related to SDS Adjustment

Recommendation 1: When there is a conceptual relationship (i.e., logical rationale or theory) between sociodemographic factors and outcomes or processes of care and empirical evidence (e.g., statistical analysis) that sociodemographic factors affect an outcome or process of care reflected in a performance measure:

- those sociodemographic factors should be included in risk adjustment of the performance score (using accepted guidelines for selecting risk factors) unless there are conceptual reasons or empirical evidence indicating that adjustment is unnecessary or inappropriate;

AND

- the performance measure specifications must also include specifications for stratification of a clinically-adjusted version of the measure based on the sociodemographic factors used in risk adjustment.

Recommendation 2: NQF should define a transition period for implementation of the recommendations related to sociodemographic adjustment. During the transition period, if a performance measure is

adjusted for sociodemographic status, then it also will include specifications for a clinically-adjusted version of the measure only for purposes of comparison to the SDS-adjusted measure.

Recommendation 3: A new NQF standing committee focused on disparities should be established. A standing disparities committee would review implementation of the revised policy about sociodemographic adjustment as recommended in this report (including key decisions by developers and purchasers) and monitor for any unintended consequences of the revised policy.

Recommendation 4: The NQF criteria for endorsing performance measures used in **accountability applications** (e.g., public reporting, pay-for-performance) should be revised as follows to indicate that patient factors for risk adjustment include <u>both</u> clinical and sociodemographic factors:

> ***2b4.*** *For outcome measures and other measures when indicated (e.g., resource use, some process):*
> *an evidence-based risk-adjustment strategy ~~(e.g., risk models, risk stratification)~~ is specified; is based on patient factors (<u>including clinical and sociodemographic factors</u>) that influence the measured outcome ~~(but not factors related to disparities in care or the quality of care)~~ and are present at start of care;[14,~~15~~] and has demonstrated adequate discrimination and calibration **OR** rationale/data support no risk adjustment~~/ stratification~~.*
> *14. Risk factors that influence outcomes should not be specified as exclusions.*
> ~~*15. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.*~~

Recommendation 5: The same guidelines for selecting clinical and health status risk factors for adjustment of performance measures may be applied to sociodemographic factors, and include the following:

- Clinical/conceptual relationship with the outcome of interest
- Empirical association with the outcome of interest
- Variation in prevalence of the factor across the measured entities
- Present at the start of care
- Is not an indicator or characteristic of the care provided (e.g., treatments, expertise of staff)
- Resistant to manipulation or gaming
- Accurate data that can be reliably and feasibly captured
- Contribution of unique variation in the outcome (i.e., not redundant)
- Potentially, improvement of the risk model (e.g., risk model metrics of discrimination, calibration)
- Potentially, face validity and acceptability

Recommendation 6: When there is a conceptual relationship and evidence that sociodemographic factors affect an outcome or process of care reflected in a performance measure submitted to NQF for endorsement, the following information should be included in the submission:

- A detailed discussion of the rationale and decisions for selecting or not selecting sociodemographic risk factors and methods of adjustment (including a conceptual description of

relationship to the outcome or process; empirical analyses; and limitations of available sociodemographic data and/or potential proxy data) should be submitted to demonstrate that adjustment incorporates relevant sociodemographic factors unless there are conceptual reasons or empirical evidence indicating that adjustment is unnecessary or inappropriate.
- In addition to identifying current and planned use of the performance measure, a discussion of the limitations and risks for misuse of the specified performance measure.

Recommendations Relevant to NQF Policy

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

- guidance for each measure as part of the endorsement process;
- guidance for different accountability applications (e.g., use in pay-for-performance versus pay-for-improvement; innovative approaches to quality measurement explicitly designed to reduce disparities).

Recommendation 8: NQF should make explicit the existing policy that endorsement of a performance measure is for a specific context as specified and tested for a specific patient population (e.g., diagnosis, age), data source (e.g., claims, chart abstraction), care setting (e.g., hospital, ambulatory care), and level of analysis (e.g., health plan, facility, individual clinician). Endorsement should not be extended to expanded specifications without review and usually additional testing.

Recommendations about Broader Related Policy Issues

Recommendation 9: When performance measures are used for accountability applications such as public reporting and pay-for-performance, then purchasers, policymakers and other users of performance measures should assess the potential impact on disadvantaged patient populations and the providers/health plans serving them to identify unintended consequences and to ensure alignment with program and policy goals. Additional actions such as creating peer groups for comparison purposes could be applied.

Recommendation 10: NQF and others such as CMS, Office of the National Coordinator (ONC) for Health Information Technology, and the Agency for Healthcare Research and Quality (AHRQ) should develop strategies to identify a standard set of sociodemographic variables (patient and community-level) to be collected and made available for performance measurement and identifying disparities.

Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors

TECHNICAL REPORT

Section 1: Introduction

NQF endorses performance measures that are suitable for both performance improvement and "accountability applications" (e.g., pay-for-performance, public reporting), when those measures meet a standard set of criteria. Measures of outcomes of care are among those endorsed by NQF. Clinical outcomes (e.g., survival, improvement or maintenance of function, relief of pain or distressing symptoms) are considered important for performance measurement because they often are the reasons for seeking and providing healthcare and reflect the quality of care received. Other outcomes for which measures may be endorsed include cost or resource use, referred to broadly as economic outcomes.

Because outcomes can be influenced by many factors other than the healthcare services and interventions received, the current NQF criteria include risk adjustment or stratification for outcome performance measures on the basis of clinical factors like comorbidity or severity of illness. In general, more severe or more complex disease in a cohort of patients, all else being equal, is associated with poorer outcomes. Risk adjustment is designed to improve the ability to make comparative conclusions about quality. Avoiding incorrect conclusions or inferences about quality is important to consumers/patients and purchasers in making informed decisions about where to obtain care; to payers, health plans, and providers regarding rewards/penalties; and to providers and plans in terms of reputation and the ability to improve care for the various subpopulations that they serve.

Current NQF criteria for performance measures direct that some sociodemographic factors, for which disparities in quality of care have been documented in the past, such as socioeconomic status (SES) and race, should not be included in statistical risk models; the related current NQF guidance (provided in a footnote) indicates that stratification is the preferred approach for these factors. The main reason for this current position on sociodemographic factors was a concern that adjustment for variables like income, education, or English proficiency would "mask disparities," and essentially allow or create lower standards of performance for "disadvantaged"[a] populations. The current criterion and concern are examined in this report.

Risk adjusting outcome performance measures to account for differences in patient health status and clinical factors (e.g., co-morbidities, severity of illness) that are present at the start of care is widely accepted. This report explores also adjusting performance measures for sociodemographic status (SDS) when appropriate. See Box 1 for examples of clinical and sociodemographic factors that affect complexity of condition, which can influence patient outcomes.

[a] In this report, "disadvantaged" is used to refer to social, economic, and/or environmental disadvantage. It could be related to a variety of sociodemographic factors such as income, race, and education.

NQF also endorses process performance measures, which typically are not adjusted for clinical or SDS. SDS adjustment of process performance measures also will be addressed in this report.

Box 1. Clinical and Sociodemographic Complexity

Clinically Complex Patient	Sociodemographically Complex Patient
• Multiple Chronic Conditions • Severe Primary Condition (e.g., severe heart failure, metastatic cancer, end-stage renal disease) • Concurrent mental and physical health problems • Disease affects multiple organ systems • Disease causes significant functional deficit or disability • Condition requires treatment by multiple providers and/or specialized sites of care	• Poverty – Low income and/or no liquid assets • Low levels of formal education, literacy, or health literacy • Limited English proficiency • Minimal or no social support –not married, living alone, no help available for essential health-related tasks • Poor living conditions – homeless, no heat or air conditioning in home or apartment, unsanitary home environment, high risk of crime • No community resources – social support programs, public transportation, retail outlets

Reason to Re-examine the NQF Policy

The increased use of NQF-endorsed performance measures beyond public reporting and quality improvement to other accountability applications, such as payment rewards and penalties, has brought increased scrutiny to performance measures. The validity and fairness of some performance measures that do not account for patients' sociodemographic complexity used to make comparative conclusions have been questioned. Consequently, reaching consensus on NQF endorsement of outcome performance measures for use in accountability applications has become increasingly controversial over the issue of adjusting outcome performance measures for SES or other sociodemographic factors. Recent examples are NQF# 1789: Hospital-wide all-cause unplanned readmission (See the Readmissions Project, section titled Candidate Consensus Standards Review) and NQF# 2158-Medicare Spending per Beneficiary Measure (MSBP) (See Cost and Resource Use Phase 1, section titled Pre-Meeting Member Comment, Phase 1).

The impact of sociodemographic factors on health and healthcare has been well documented.[1-3] In fact, most epidemiological and health services research studies that focus on quality commonly adjust for patient SES. In contrast, SES adjustment of quality measures has been typically avoided. There are at least two divergent views regarding adjustment for sociodemographic factors:

1) Adjusting performance measures for sociodemographic factors is essential to making fair comparative conclusions about quality and is important to consumers/patients, payers, and others making decisions about choice of providers or health plans or assigning rewards or penalties. Disadvantaged patients confront varying barriers, often lifelong, to health and healthcare, and failing to account for the sociodemographic factors when indicated creates an uneven playing field for performance measurement. For example, Satin[4] states "Asking clinics and physicians who work primarily with poor patient populations to achieve the same results as those working with wealthier populations is effectively asking for more, and in some cases, impossibly more from these providers/plans. The results of such unrealistic demands may be fewer and fewer providers/plans willing to serve the already underserved."
2) Adjusting performance measures for sociodemographic factors should not be done because it obscures disparities and implies that differences in outcomes based on SDS are expected and accepted. For example, Iezzoni[5, p. 21] states: "For some purposes, ethical concerns raise questions

about whether and how to risk-adjust. Such situations arise when persons with certain attributes (e.g., gender, race, SES) that might be potential risk factors for a given outcome simultaneously face the likelihood of receiving substandard care because of those attributes."

Interestingly, both of these positions are based in part on a shared concern about entrenching or worsening disparities in health or healthcare. In the first view, if performance measurement fails to recognize sociodemographic complexity, then it may create a disincentive for healthcare providers and health plans to serve disadvantaged patients, decreasing access to healthcare. In the second view, if performance measurement adjusts for sociodemographic factors, then it may create a disincentive for healthcare providers and plans to improve care to disadvantaged patients.

The issues and concerns about the potential unintended consequence of adjusting or not adjusting for sociodemographic factors on disparities for "disadvantaged" patient populations are addressed in more detail later. However, it is important to note that any recommendations about risk adjusting performance measures must be grounded in sound measurement science, which also is addressed in this report.

Terminology and Key Definitions

In this report, the following key terms are used.

- **Unit** will be used to signify the entity whose performance is being measured, which could be a hospital, health plan, clinician, etc. Performance measurement (and sociodemographic adjustment) can be applied to any setting and level of analysis.
- **Clinical adjustment** refers to adjustment for only clinical variables.
- **Sociodemographic or SDS adjustment** refers to adjustment for both clinical and sociodemographic variables.

The key concepts used in this report are defined as follows and also included in the glossary in Appendix B.

- **Confounding** refers to the distortion in the degree of association between an exposure (independent variable) and an outcome (dependent variable) due to a mixing of effects between the exposure and an incidental (confounding) factor. Confounding represents systematic error and threatens the internal validity of an epidemiologic study since it can lead to false conclusions regarding the true relationship between an exposure and outcome. (See the basics of confounding in Appendix D.)

- **Risk adjustment (also known as case-mix adjustment)** refers to statistical methods to control or account for patient-related factors when computing performance measure scores; methods include multivariable modeling, indirect standardization, or direct standardization. These methods can be used to produce a ratio of observed-to-expected, a risk-adjusted rate, or other estimate of performance. (See the basics of risk adjustment in Appendix C.)

- **Stratification** refers to computing performance scores separately for different strata or groupings of patients based on some characteristics(s)—i.e., each healthcare unit has multiple performance scores (one for each stratum) rather than one overall performance score.

- **Peer groups for comparison** refers to creating peer groups of healthcare units caring for a similar mix of patients, within which to examine performance scores.

- **Sociodemographic Status (SDS)** refers to a variety of socioeconomic (e.g., income, education, occupation) and demographic factors (e.g., race, ethnicity, primary language).

- **Outcome** – the result of providing healthcare. The term, outcome, will be used to broadly include the following types of outcomes relevant to performance measurement: **quality outcomes** of health outcome (e.g., mortality), intermediate clinical outcome (e.g., BP < 140/90), patient-reported outcome (e.g., depression), and **economic outcomes** of cost and resource use.

Project Purpose, Scope, Approach

There is a large body of evidence that various sociodemographic factors influence outcomes, and thus influence results on outcome performance measures. There also is a large body of evidence that there are disparities in health and healthcare related to some of those sociodemographic factors. Given the evidence, the overarching question addressed in this project is "What, if anything should be done about sociodemographic factors in relation to outcome performance measurement?"

The purpose of this project was to:

- Identify and examine the issues related to risk adjusting outcome performance measures for SDS (i.e., SES and/or other sociodemographic factors).
- Make recommendations regarding if, when, for what, and how outcome performance measures should be adjusted for SES or other sociodemographic factors.
- Make recommendations for NQF's endorsement criteria for outcome performance measures.

During the project, the Expert Panel identified that process performance measures also may need adjustment.

This project did not include recommendations for:

- specific performance measures;
- adjustment for determining payment for services provided, such as capitated payments;
- use of a particular risk adjustment or statistical procedures; or
- structuring performance reward/penalty programs such as pay-for-performance.

A multistakeholder Expert Panel (Appendix A) with a variety of experiences related to outcome performance measurement and disparities reviewed the issues and made recommendations regarding the use of SES and other sociodemographic variables for adjusting outcome performance measures. The Expert Panel's draft recommendations were presented for public comment. This report and the recommendations reflect the Expert Panel's modifications in response to comments.

Core Principles

The Expert Panel agreed on a set of core principles to ground its recommendations. The principles were not intended to imply a particular direction for recommendations related to risk adjustment for SES and

sociodemographic factors; rather, they represented a baseline of agreement on the key issues that must be considered in making recommendations.

1. Outcome performance measurement is critical to the aims of the National Quality Strategy.
2. Performance measurement and risk adjustment must be based on sound measurement science.
3. Disparities in health and healthcare should be identified and reduced.
4. Performance measurement should not lead to increased disparities in health and healthcare.
5. 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.
6. When used in accountability applications, performance measures that are influenced by factors other than the care received, particularly outcomes, need to be adjusted for relevant differences in patient case mix to avoid incorrect inferences about performance.
7. Risk adjustment may be constrained by data limitations and data collection burden.
8. The methods, factors, and rationale for risk adjustment should be transparent.

Section 2: Recommendations

The Expert Panel made the following ten recommendations. A brief rationale accompanies each recommendation in this section. However, an in-depth discussion of the methodological basis and other considerations that led the Panel to these recommendations is in the following sections.

Although the draft recommendations were supported by the great majority of the Expert Panel and the NQF member and public commenters, the purchaser stakeholders and some, but not all, of the consumer stakeholders remained concerned about appropriateness of adjusting for SDS. The Expert Panel carefully considered these ongoing concerns and modified their draft recommendations in three substantial ways:

- requiring measure specifications for stratification to identify disparities if a performance measure is SDS-adjusted;
- recommending a transition period during which a clinically-adjusted version of the performance measure would be specified and available only for comparison purposes to the SDS-adjusted score; and
- recommending an NQF Standing Disparities Committee to monitor implementation of the revised policy as well as ensure continuing attention to disparities.

In addition, the Expert Panel provided a more detailed methodological discussion (Section 4) to facilitate better understanding of what risk adjustment does and does not do. See Appendix G for comment themes and Panel responses.

Applicability of Recommendations

The recommendations may apply to outcome performance measures (including resource use and patient-reported outcomes) and some process performance measures used for comparative performance assessment. However, ***each performance measure must be assessed individually to determine appropriateness of sociodemographic adjustment***. The recommendations may apply to any level of analysis including health plans, facilities, individual clinicians, accountable care organizations, etc.

Recommendations Related to NQF Criteria and Processes Related to SDS Adjustment

Recommendation 1: When there is a conceptual relationship (i.e., logical rationale or theory) between sociodemographic factors and outcomes or processes of care and empirical evidence (e.g., statistical analysis) that sociodemographic factors affect an outcome or process of care reflected in a performance measure:

- those sociodemographic factors should be included in risk adjustment of the performance score (using accepted guidelines for selecting risk factors) unless there are conceptual reasons or empirical evidence indicating that adjustment is unnecessary or inappropriate;

AND

- the performance measure specifications must also include specifications for stratification of a clinically-adjusted version of the measure based on the sociodemographic factors used in risk adjustment.

Rationale: Patient characteristics that are present before care begins can influence patient outcomes or some processes of care. In order to avoid incorrect inferences (or conclusions) about quality in the context of comparative performance evaluation of various healthcare entities, some performance measures need to be adjusted for relevant patient characteristics when certain conditions are met.

Adjustment of performance measures for clinical complexity of the mix of patients is widely accepted and the same principles and methods may apply to sociodemographic characteristics. There are conceptual and statistical conditions for selecting risk factors that must be met and evaluated for each individual performance measure. Not all performance measures, or even all outcome performance measures, may need to be adjusted for sociodemographic factors. For example, the outcome of central line infection occurring during a hospital stay or the process of administering the correct medication at the correct time during a procedure would not have a conceptual basis for SDS adjustment. However, if there is a conceptual (i.e., logical rationale or theory, prior research) and empirical relationship with the outcome or process being measured (i.e., based on statistical analysis) and the guidance for selecting risk factors is followed, relevant SDS factors should be included in risk adjustment procedures to avoid incorrect inferences based on an overall performance score. This approach is grounded in accepted methods and principles related to statistical inference and confounding discussed in Section 4.

The recommendation acknowledges there may be situations where SDS adjustment is unnecessary or inappropriate based on conceptual reasons or empirical evidence. The information submitted with a performance measure considered for NQF endorsement should justify the approach taken as outlined in the recommendations. These topics are discussed in Sections 4 and 6.

Identifying and reducing disparities in health and healthcare are important national priorities and require additional analysis of performance data by patient subgroups. If sociodemographic factors are included in a risk model, it indicates that the measure is disparities-sensitive and should also be stratified to identify differences by patient subgroups. Stratified performance data are most useful and most transparent as a means of identifying where disparities exist, which isn't possible in an overall score, whether only clinically-adjusted, or SDS-adjusted. This is a continuation and strengthening of NQF's prior guidance to stratify disparities-sensitive performance measures by requiring the measure also be specified for stratification. Performance data should be stratified on the basis of the sociodemographic factors used in risk adjustment so that clinically-adjusted scores are computed for each stratum (not one overall clinically-adjusted score). Specifications would include how the strata are constructed and how to compute the clinically-adjusted score for those strata. It is important to note a major limitation of stratified data by healthcare unit: small cell sizes decrease the reliability of the estimates and they should not be used for comparative performance evaluation. Appropriate explanations about limitations or minimum cell sizes to be reported should accompany the stratified data.

Clearly, a concerted effort among providers, health plans, policymakers, researchers, and the public is needed to address healthcare disparities. For example, when sociodemographic factors influence a performance measure, providers need to examine their own data to identify opportunities for improvement in serving disadvantaged patient populations. The Centers for Medicare & Medicaid Services (CMS) or other producers of performance reporting should make such stratified data easily available to interested parties, such as consumer advocates, researchers, health plans, and providers.

Doing so could serve a dual purpose of providing finer grained data to interested parties and for assessing and addressing healthcare disparities.

Recommendation 2: NQF should define a transition period for implementation of the recommendations related to sociodemographic adjustment. During the transition period, if a performance measure is adjusted for sociodemographic status, then it also will include specifications for a clinically-adjusted version of the measure only for purposes of comparison to the SDS-adjusted measure.

Rationale: A defined transition period with specific evaluation parameters will facilitate a systematic collection of information about the change in policy, including additional information about the effects of sociodemographic adjustment and any unintended consequences. Additional guidance related to implementing stratification as outlined in recommendation 1 may need to be developed. Therefore, during the transition period, specifications for a clinically-adjusted version of the SDS-adjusted measure would be included within the SDS-adjusted measure submission and identified as endorsed for comparison purposes only. "Comparison" here means comparison between overall scores of the clinically-adjusted and SDS-adjusted versions of a measure to understand the effects of SDS adjustment. It does not mean use of the clinically-adjusted measure for actual comparisons of health plans or providers in public reporting or pay-for-performance programs. The clinically-adjusted version of the fully adjusted measure is an essential step to stratification as recommended and also has been seen by some stakeholders as important to understanding the effect of the policy change. The second part of recommendation 1 indicates that an endorsed SDS-adjusted measure always includes specifications for stratification of the clinically-adjusted version of the measure; therefore, specifying a clinically-adjusted version of the measure is a required step toward stratification. The recommended Disparities Committee would be tasked with further detailing requirements for stratification.

Recommendation 3: A new NQF standing committee focused on disparities should be established.

Rationale: A standing disparities committee would review implementation of the revised policy about sociodemographic adjustment as recommended in this report (including key decisions by developers and purchasers) and monitor for any unintended consequences of the revised policy. It would also assess trends in disparities and review and provide guidance related to methodologies for adjustment and stratification such as use of community factors, and standard sociodemographic data collection. The membership of the committee should follow standard NQF policy about representation of diverse stakeholders and balance of perspectives.

Such a committee would also help ensure that social and demographic disparities in care do not get overlooked, but rather remain an integral part of quality measurement. The committee would be explicitly tasked with examining evidence for unintended consequences to patients across the full range of NQF-endorsed measures—including lowered expectations and incentives to improve care to disadvantaged patients—by monitoring disparities both between and within providers. The committee would review decisions regarding when measures are adjusted for sociodemographic factors and how. It would assess the impact of the NQF policy changes on disadvantaged patients and on safety net providers. It would recommend the collection of additional sociodemographic data (individual- or

community-level). The committee would suggest ways to better address and/or integrate healthcare equity and value. The committee could investigate how risk adjustment methodologies and stratification may influence our understanding of where and why disparities exist. It also could play a role in assisting developers and end users understand the role of risk adjustment and stratification in portraying and evaluating provider and health plan performance.

Because of the change to long-standing NQF policy proposed in the panel's recommendations, the disparities committee would be specifically tasked with preparation of an annual report, for at least the first five years of its existence, for public release, on the issues listed above. Its first task would involve a one-year look back at the consequences of the recommendations, both intended and unintended. This would help ensure that the recommendations were having the intended effect.

Recommendation 4: The NQF criteria for endorsing performance measures used in **accountability applications** (e.g., public reporting, pay-for-performance) should be revised as follows to indicate that patient factors for risk adjustment include <u>both</u> clinical and sociodemographic factors:

> ***2b4.*** *For outcome measures and other measures when indicated (e.g., resource use, some process):*
> *an evidence-based risk-adjustment strategy ~~(e.g., risk models, risk stratification)~~ is specified; is based on patient factors (<u>including clinical and sociodemographic factors</u>) that influence the measured outcome ~~(but not factors related to disparities in care or the quality of care)~~ and are present at start of care;[14,~~15~~] and has demonstrated adequate discrimination and calibration* ***OR*** *rationale/data support no risk adjustment~~/ stratification~~.*
> *14. Risk factors that influence outcomes should not be specified as exclusions.*
> *~~15. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.~~*

Rationale: This change in the NQF criteria removes the prohibition against adjusting for sociodemographic factors and is consistent with recommendation 1.

Recommendation 5: The same guidelines for selecting clinical and health status risk factors for adjustment of performance measures may be applied to sociodemographic factors, and include the following:

- Clinical/conceptual relationship with the outcome of interest
- Empirical association with the outcome of interest
- Variation in prevalence of the factor across the measured entities
- Present at the start of care
- Is not an indicator or characteristic of the care provided (e.g., treatments, expertise of staff)
- Resistant to manipulation or gaming
- Accurate data that can be reliably and feasibly captured

- Contribution of unique variation in the outcome (i.e., not redundant)
- Potentially, improvement of the risk model (e.g., risk model metrics of discrimination, calibration)
- Potentially, face validity and acceptability

Rationale: The guidelines for selecting clinical risk factors apply equally well to sociodemographic factors. Selecting risk factors and developing a model is an iterative process, but is based first on a conceptual relationship and demonstration of an empirical relationship with the outcome or process of interest. A detailed discussion of selecting risk factors is provided in Section 6.

Recommendation 6: When there is a conceptual relationship and evidence that sociodemographic factors affect an outcome or process of care reflected in a performance measure submitted to NQF for endorsement, the following information should be included in the submission:

- A detailed discussion of the rationale and decisions for selecting or not selecting sociodemographic risk factors and methods of adjustment (including a conceptual description of relationship to the outcome or process; empirical analyses; and limitations of available sociodemographic data and/or potential proxy data) should be submitted to demonstrate that adjustment incorporates relevant sociodemographic factors unless there are conceptual reasons or empirical evidence indicating that adjustment is unnecessary or inappropriate.
- In addition to identifying current and planned use of the performance measure, a discussion of the limitations and risks for misuse of the specified performance measure.

Rationale: NQF submission currently requires information on risk adjustment specifications, risk factor selection, assessment of the risk adjustment procedure, and current and planned use of the performance measure. The developer's decisions regarding sociodemographic factors, including use of proxy data, should be transparent and open to review and evaluation.

Recommendations Relevant to NQF Policy

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

- guidance for each measure as part of the endorsement process;
- guidance for different accountability applications (e.g., use in pay-for-performance versus pay-for-improvement; innovative approaches to quality measurement explicitly designed to reduce disparities).

Rationale: A measure that is ideal for one use may not be ideal for another. 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 review of the detailed information about the performance measure for potential endorsement provides an opportunity to identify any specific considerations or limitations for use in specific accountability applications.

Recommendation 8: NQF should make explicit the existing policy that endorsement of a performance measure is for a specific context as specified and tested for a specific patient population (e.g., diagnosis, age), data source (e.g., claims, chart abstraction), care setting (e.g., hospital, ambulatory care), and level of analysis (e.g., health plan, facility, individual clinician). Endorsement should not be extended to expanded specifications without review and usually additional testing.

Rationale: This is implicit in the current NQF criteria and process for endorsing a performance 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 would require an ad hoc review to expand endorsement.

Recommendations about Broader Related Policy Issues

Recommendation 9: When performance measures are used for accountability applications such as public reporting and pay-for-performance, then purchasers, policymakers and other users of performance measures should assess the potential impact on disadvantaged patient populations and the providers/health plans serving them to identify unintended consequences and to ensure alignment with program and policy goals. Additional actions such as creating peer groups for comparison purposes could be applied.

Rationale: Even if a performance measure is adjusted using sociodemographic factors, this does not ensure protection of safety net providers and additional strategies may be needed. For example, SDS adjustment or stratification for patient-level factors does not address potential differences in community factors such as public funding or area healthcare resources, which may have a substantial impact on comparative performance results. Given that safety net providers are differentially funded (a function of local and state taxing jurisdictions), making comparisons even among safety net providers may be problematic. Accountability programs should consider if and how to incorporate this type of community factor into comparative evaluations for purposes of assigning rewards and penalties.

Although NQF does not control how measures are implemented, it is important to signal that the impact of program polices on providers or health plans caring for disadvantaged populations should be considered. These units may have fewer resources to improve the care they provide. The recent MedPAC recommendation regarding hospital readmissions is an example of creating peer groups for comparison as a way to lessen the impact of a performance penalty on safety-net hospitals.

Recommendation 10: NQF and others such as CMS, Office of the National Coordinator (ONC) for Health Information Technology, and the Agency for Healthcare Research and Quality (AHRQ) should develop strategies to identify a standard set of sociodemographic variables (patient and community-level) to be collected and made available for performance measurement and identifying disparities.

Rationale: Even when performance measures should be adjusted for sociodemographic factors, data limitations currently pose a substantial barrier. Although mandated data collection is beyond the scope

of NQF, there is a need for a national effort to collect relevant sociodemographic information in a standardized way that allows for its valid use in adjustment models that will be applied across states and regions. Most sociodemographic variables, particularly socioeconomic factors, that could conceivably be used in risk adjustment models are not currently collected in a standard way by health plans, doctors, hospitals, and other healthcare providers, and are not included in claims data bases that are often used to develop risk models. Data on sociodemographic factors also are important for providers when providing care and when reviewing their performance for quality improvement.

Section 3: Background

Context of Comparative Performance Assessment

NQF endorses performance measures that are intended for use in accountability applications such as public reporting and pay-for-performance. In this context, the overall performance measure score is used to make a conclusion about a unit's quality in relation to other units or some other comparator such as average performance. The general question being addressed is: ***how would the performance of various units compare if hypothetically they had the same mix of patients*****?** That is, the measure scores are used to identify which units have better quality in order to inform decisions of an individual to seek care, a purchaser to pay for care or give a bonus or penalty, for networks to contract, etc. Such comparisons should be affected as little as possible by factors other than quality of care, including patient characteristics.

Because healthcare outcomes are a function of patient attributes (including SDS) as well as the care received; and patients are not randomly assigned to units for healthcare services so that all have the same mix of patients, risk adjustment is essential to examining outcome performance in real-world settings.[5] Thus, when comparing outcomes, the purpose of risk adjustment is to ensure like-to-like comparisons. [5] Without appropriate risk adjustment, units can be misclassified based on incorrect conclusions about comparative performance. (See the basics of risk adjustment in Appendix C.) Depending on the specific program in which the performance measures are used, misclassification can create disincentives to care for more complex patients (clinically or sociodemographically complex) and potentially decrease resources to those units with large shares of complex patients.

Although NQF does not control the structure of various accountability programs, NQF's primary role is to ensure that an endorsed performance measure is suitable for use in comparative accountability applications. An appropriately adjusted performance measure alone will not solve other issues or problems that could be present in various accountability programs or formulas for determining base payment for services to more complex patients, which are outside the role of NQF.

Evidence-Based Risk Adjustment Strategy

NQF measure evaluation criteria call for an "evidence-based" risk adjustment strategy. Identifying potential risk factors may be informed by prior studies, but it is not required. Ultimately the final risk adjustment strategy requires empirical evidence from the statistical analyses regarding the relationship of the potential factors to the outcome, first individually and then in the context of other risk factors. Risk factors and their strength of association are unique to each individual performance measure. The requirement for an evidence-based risk adjustment strategy is different from the NQF requirement for clinical evidence that supports performance measures of structure, processes, and intermediate outcomes and calls for a systematic assessment and grading of the body of clinical evidence that supports their link to desired outcomes.

Sociodemographic Factors and Outcomes

The term *sociodemographic* will be used to include a variety of socioeconomic (e.g., income, education, occupation) and demographic factors (e.g., race, ethnicity, primary language) that are often associated with disadvantage among affected populations. Although age is a demographic factor, it also is considered a clinical factor and already included in many risk adjustment procedures. A large body of evidence shows an association between various sociodemographic variables and outcomes.[1-3] In

general, sociodemographic "disadvantage" (e.g., low income, low education, homelessness) is often associated with poorer patient outcomes (e.g., higher morbidity, mortality, or readmissions). Low SES and social disadvantage tend to be associated with greater morbidity, disease severity, and worse quality of life.[6, 7]

The mechanism(s) for the association between sociodemographic factors and health status and outcomes is often complex and is not always be clear.[8, 9] Depending on the specific SDS factor and outcome, it can involve the effect of mediators such as financial resources, community resources, or patient understanding on the ability to access healthcare services or follow through with treatments. These factors contribute to healthcare disparities.[10] Sociodemographic factors operate in the present but also may have a cumulative effect on health outcomes across the life course through a variety of mechanisms including early effects on sensitive periods during development and epigenetic effects. Historical and current discrimination impact the patient, ranging from biological stress levels to social confidence when interacting with the healthcare system. Another potential and simultaneous mechanism may be the implicit biases or assumptions on the part of healthcare providers that influence their interactions with, and the care options given to, patients with different characteristics (e.g., race/ethnicity), thus increasing the likelihood of providing substandard care. Disadvantaged patients also may be concentrated in areas of poorly resourced or lower quality healthcare services.

The characteristics associated with being disadvantaged (e.g., low SES) generally are associated with less than optimal clinical outcomes. However, for resource use and cost outcomes, the relationship could vary. Depending on timing and the population included, cost and resource use could be less in disadvantaged patients because of inability to access preventive and early diagnostic services.

Essentially, the evidence of a relationship to SDS will vary depending on the specific outcome or process being measured. As will be discussed in the report, potential risk factors need to be assessed empirically with actual data for the proposed risk factors and the outcome being measured.

Process Performance Measures

Most of the same issues regarding the relationship between sociodemographic factors and outcomes might also apply to processes, especially processes that are not directly under the control of the healthcare provider and require some action by the patient (e.g., getting prescription filled). As with the outcome performance measures, adjusting process measures should be guided first and foremost by a causal theory. Many processes are primarily under the control of healthcare providers (e.g., administering the correct antibiotic to prevent surgical site infection), and adjustment for sociodemographic factors would not be appropriate because the relevant clinical guideline generally would make no exception for sociodemographic factors, and there is no plausible, acceptable causal path through which a sociodemographic factor would affect performance of the clinical process. Some processes, though, are not as strongly under the control of the provider (e.g., adherence with medications, receipt of screening colonoscopy), and adjustment for sociodemographic factors might be called for if the general criteria for selecting risk factors are met. The recommendations regarding sociodemographic adjustment also apply to some process performance measures.

Perspectives on Adjusting for Sociodemographic Factors

The reasons for and against adjusting performance measures for SDS were identified during the Panel's deliberations and they also were raised during the comment period. The reasons for opposing or

supporting SDS adjustment are based on a combination of evidence, logical arguments, and some assumptions about drivers of behavior.

The Expert Panel carefully considered all perspectives and ultimately recommended that performance measures be adjusted for SDS under certain conditions. The concerns about negative consequences attributed to SDS adjustment are specifically addressed in the recommendations to include specifications to stratify any performance measures that are SDS-adjusted, establish a Standing Disparities Committee, and define a transition period. The concerns are also addressed in the methodological discussion.

During the review and comment process, some questioned the role of evidence in weighing the arguments for and against SDS adjustment. Ultimately, the recommendation for SDS adjustment could only be made if it was based on sound measurement science, which will be discussed in this report.

Concerns and Unintended Consequences about Adjusting Performance Measures for Sociodemographic Factors

The first and most important concern about adjustment for sociodemographic factors is that disadvantaged patient groups, on average, might receive worse quality of care. In other words, differences in observed performance, either across units or by patient group within units, reflect actual differences in the processes of care for disadvantaged versus other patients that would be "adjusted away". A systematic review of quality of care related to SES, Asch et al.[11] found small, but statistically significant differences in quality of care provided by the income level of patients. Unexpectedly, Blacks and Hispanics received slightly better quality care than Whites. However, this study did not examine potential mechanisms for worse care for poorer patients.

There are three mechanisms through which providers might provide worse care processes for disadvantaged patients. First, it could reflect bias in care by providers in general based on the sociodemographic characteristics of the patients (e.g., poverty, race, language).[12, 13] Second, it could reflect reduced resources and funding in places where patients receive care. If disadvantaged patients cluster within poorly resourced units or within units that provide worse care, then disadvantaged patients will on average, receive worse care.[14-16] A third mechanism involves attempts by the clinician to tailor care to perceived constraints by the patient. Such decisions might be appropriate, i.e., when they are collaboratively made in partnership with the patient, or inappropriate, i.e., when the physician unilaterally decides what the patient wants and/or can afford. Findings from empirical studies often differ depending on the performance measure and provider type.

A second concern is that adjustment will mask meaningful differences in quality or performance—that is, the adjustment will have a strong enough effect that meaningful differences in performance will not be detectable in adjusted performance scores. The concern reflects a belief that differences in unit outcome performance reflect the degree to which units implement interventions to mitigate the effects that sociodemographic factors have on those outcomes (e.g., instructions in multiple languages, interpreters, prescribing low-cost generic drugs, hospital discharge follow-up), rather than the effect of those factors on patient outcomes.

A third concern is that adjustment implies that worse outcomes are "expected" for certain patient groups such as those with low income, creating a double standard, and no expectation that providers try to mitigate the effect of such factors on outcomes. Some worry that if adjusting for sociodemographic

differences results in performance being labeled "average" or "as expected" despite worse outcomes for disadvantaged patients, it will blunt the motivation to provide optimal care for disadvantaged patients. In other words, if the effect of sociodemographic factors is "adjusted away," one cannot or will not do something about them. Empirical evidence supporting or refuting this concern is lacking.

Concerns and Unintended Consequences When Performance Measures are Not Adjusted for Sociodemographic Factors

The association between SDS and outcomes has also been demonstrated with outcome performance measures for physicians, hospitals, and other healthcare providers.[9, 16-21] In general, caring for sociodemographically "disadvantaged" populations is associated with poorer performance (based on current performance measures) on average, although there are some noteworthy exceptions to the general pattern.[22]

As discussed in the methodological basis for adjustment, SDS is a potential confounder and can lead to incorrect comparative inferences or conclusions about quality. An alternative explanation to the conclusion of poor quality for poorer performance scores is that the unit is caring for a disproportionate share of disadvantaged patients, who all else being equal, have worse outcomes (just as do clinically complex patients).

In addition to hindering informed decisionmaking by patients, use of performance measures that fail to account for sociodemographic factors when indicated, could lead to harm of patients through other mechanisms. As healthcare moves toward increasing use of financial rewards for better quality and financial penalties for worse quality, use of measures that result in incorrect conclusions about quality poses a substantial risk for penalizing healthcare organizations and providers who serve more disadvantaged populations.[23-27]

Units serving "low-sociodemographic" populations and communities are more likely to be identified as "poor performers" and either be less likely to receive financial rewards, or be more likely to face financial penalties, in pay-for-performance programs. Joynt and Jha,[28] for example, found that safety-net hospitals were more than twice as likely as other hospitals to have high penalties in the first year of the Medicare Hospital Readmission Reduction Program. In another example, Young, et al., found a strong association between socioeconomic characteristics of members of Medicare Part D drug plans and the performance ratings of those plans[29, 30]

In the context of public quality reporting and pay-for-performance, failing to account for the greater difficulty in achieving good outcomes in socially and economically disadvantaged populations could set up a series of adverse feedback loops that result in a "downward spiral" of access and quality for those populations. The net effect could worsen rather than ameliorate healthcare disparities. There are at least three potential adverse consequences, each of which could have the eventual effect of undermining the quality of care for disadvantaged patients, thus exacerbating disparities in health and healthcare.

First, with public reporting of performance, healthcare units will have a strong incentive to avoid serving disadvantaged populations, so as to avoid being labeled as a "bad performer." This could happen based on where physicians and other individual providers choose to work, where facilities are opened or closed, or expanded or contracted, where health plans operate, and through more subtle ways of

"cherry picking." A study on public reporting of surgeon mortality rates for coronary artery bypass graft (CABG) that did not adjust for race resulted in decreased access to surgery for racial/ethnic minorities.[31]

Second, with some pay-for-performance programs, substantial funding will be shifted away from organizations serving disadvantaged populations and communities and to organizations serving more affluent, less vulnerable, populations and communities.

Third, individual consumers, private and public payers, and others choosing among providers or plans whose performance will be publicly reported will tend to avoid units serving disadvantaged patients and communities,[32] based on performance scores that may not provide a valid comparative performance assessment.

In the context that measurement science supports adjustment for sociodemographic factors when certain conditions are met, most of the NQF Expert Panel members were concerned that the potential negative impact on care to disadvantaged patients resulting from the three phenomena listed above were, on balance, even more detrimental than the concerns about "masking disparities" or "setting a lower bar for performance." Therefore, the current position against sociodemographic adjustment should be reconsidered. However, the expressed concerns about masking disparities and lower standards were also addressed in the final recommendations.

Mitigation of Effect of Sociodemographic Factors

Adjustment for SDS does not mean that providers cannot take steps to mitigate the effects of some sociodemographic factors. Just as care is adjusted based on clinical severity and complexity, care should be adjusted to address specific needs related to sociodemographic factors. Some examples include providing interpreters, instructions in different languages, discharge clinics, prescribing generic drugs, outreach to homeless patients in community settings, etc. Strategies to mitigate the effects of sociodemographic factors are often resource-intensive and raise payment policy issues that are outside the scope of this project but discussed briefly later in this report.

Section 4: Methodological Basis for Risk Adjustment

As already mentioned, when performance measures are used for comparative assessments as with pubic reporting and pay-for-performance, risk adjustment is essential to avoid making incorrect inferences or conclusions about quality or performance. The goal of risk adjustment is to be able to answer the question: ***how would the performance of various units compare if hypothetically they had the same mix of patients*****?**

The need for risk adjustment is based on accepted and foundational statistical theory and epidemiologic principles involving causal inference and confounding. This section describes the key foundation for risk adjustment and also provides responses to some of the concerns about sociodemographic status (SDS) adjustment from a methodological perspective. Other conceptual considerations are discussed in Section 6.

Key Definitions

- **Confounding** refers to the distortion in the degree of association between an exposure (independent variable) and an outcome (dependent variable) due to a mixing of effects between the exposure and an incidental (confounding) factor. Confounding represents systematic error and threatens the internal validity of an epidemiologic study since it can lead to false conclusions regarding the true relationship between an exposure and outcome. (See the basics of confounding in Appendix D.)
- **Risk adjustment (also known as case-mix adjustment)** refers to statistical methods to control or account for patient-related factors when computing performance measure scores; methods include multivariable modeling, indirect standardization, or direct standardization. These methods can be used to produce a ratio of observed to expected, a risk-adjusted rate, or other estimate of performance. (See the basics of risk adjustment in Appendix C.)

Terms

- **Unit** will be used to signify the entity whose performance is being measured, which could be a hospital, health plan, clinician, etc.
- **Clinical adjustment** refers to adjustment for only clinical variables.
- **Sociodemographic or SDS adjustment** refers to adjustment for both clinical and sociodemographic variables.

Conceptual Basis for Risk Adjustment

In clinical comparative effectiveness studies, researchers often ask whether one treatment is better than another for reducing morbidity or improving survival. For example, in a randomized controlled trial comparing the success rate of two treatments, say "A" and "B", the average "effect" of treatment A versus B can be estimated by the difference in the proportion of patients receiving treatment A who have a successful outcome and the proportion of patients receiving treatment B who have a successful outcome, $\hat{s}_A - \hat{s}_B$. Randomization ensures that patients receiving the two treatments are comparable. When treatments are not randomly assigned— as in a nonrandomized observational study— the observed difference $\hat{s}_A - \hat{s}_B$ may be biased. To the extent that patients in each treatment group differ in ways that affect outcomes (e.g., they are sicker, frailer, etc.), the observed differences in outcomes may reflect different patient characteristics rather than the treatment effect of interest. In other words, the effect of treatment is confounded by differences in pretreatment patient factors. An identical

confounding issue arises in studies comparing outcomes of healthcare units, in which patients are not randomized to units (just as in the real-world environment of performance measurement in which patients are not randomized to units).

Table 1 illustrates potential confounding in performance measurement. In this example, clinical severity is associated with mortality in the national patient population with a difference of one percentage point between patients with low to average clinical severity vs. high severity (2% vs. 3%). Across the units, the proportion of high severity patients varies—unit A has exactly the same proportion as nationally (20%) and unit B has a higher proportion (60%). Unit B has also a higher unadjusted mortality rate than the national average (2.6% vs. 2.2%). Because severity is a potential confounder, an alternative explanation for unit B's higher overall mortality rate is its substantially higher proportion of high-severity patients, rather than it delivers worse quality of care. The data by severity group indicate that for unit B, the higher overall mortality is a function of serving a larger proportion of higher severity patients (because the mortality within each group is exactly the same as the national averages for those groups).

Table 1. Example of Confounding

	All Patients in National Population		**Unit A**		**Unit B**	
Clinical Category	Pt mix N/Percent	**Unadjusted Mortality** N/Percent	n	**Unadjusted Mortality** n/Percent	n	**Unadjusted Mortality** n/Percent
All patients	1,000,000 100%	22,000 **2.2%**	1000	22 **2.2%**	1000	26 **2.6%**
Low-Average Severity	800,000 80%	16,000 2%	800 80%	16 2%	400 40%	8 2%
High Severity	200,000 20%	6,000 3%	200 20%	6 3%	600 60%	18 3%

Risk adjustment refers to a collection of techniques for reducing the effect of confounding factors in studies where patients are not randomly assigned to different treatments. In performance evaluation, the "treatments" are different healthcare units. The "treatment effect" may be conceived as the difference between a patient's actual outcome and the outcome that would have occurred had the patient been treated by another unit. Risk adjustment aims to control for patient factors (e.g., morbidity or sociodemographic factors) that could affect the outcome so that residual differences in outcomes reflect the treatment effect of interest.[33]

The statistical and epidemiologic literature describes conditions in which valid inferences about treatment effects based on observational data are possible. In general, unbiased estimation (i.e., without systematic deviation from the true value) requires the assumption that outcome differences are unconfounded conditional on a set of pretreatment covariates. This unconfoundedness assumption means that blocks of patients having identical values of pretreatment covariates and who receive each treatment, are like a random sample from a common population. Although the unconfoundedness assumption is unlikely to be literally true in a nonrandomized observational study, the risk of encountering large violations of the assumption can be minimized by careful consideration of all

potential confounders. The assumption becomes more plausible if the set of covariates is expanded to include all factors that may predict the outcome or the choice of healthcare unit or both. Data availability may be a practical constraint on the factors that can be considered.

What types of variables are appropriate for risk adjustment?

Risk adjustment involves an attempt to compare only patients who are alike with respect to a set of pretreatment covariates. When we say that a variable was "adjusted" or "included in the risk adjustment," we mean that the analysis aimed to compare outcomes of patients at different healthcare units who were similar with respect to that covariate. In general, covariates appropriate for risk adjustment are those factors that are hypothesized to remain the same if the patient were to be reassigned to a differentunit.[34] By this rule, any patient characteristic that is present prior to treatment and is a known or suspected confounder of the treatment effect may be included. Variables of sociodemographic complexity could also cause confounding in the same way as severity of illness in the example in Table 1.

Although it is generally desirable to adjust for all important confounding factors, theory dictates that we should not adjust for components of the treatment being evaluated.[35] Doing so may "adjust away" differences in outcomes that result from the adoption of more or less effective care practices by different units. For example, one would generally not adjust for the frequency of hand washing when comparing infection rates across hospitals because assiduous hand washing is one of the ways in which a hospital may seek to achieve a lower infection rate.[34] SDS factors are not treatment variables in the way that whether a specific surgical intervention was provided is a treatment variable, and therefore, they do not "adjust away" treatment effects. However, additional concerns about how SDS factors may influence treatment and "adjust away" the unit treatment effect are discussed in later sections below.

Does adjusting for sociodemographic factors mask disparities in outcomes for disadvantaged patients?

In terms of revealing disparities, a single SDS-adjusted score is no different than a single clinically-adjusted score. SDS adjustment may change the unit score if the proportion of disadvantaged patients is larger or smaller than average. However, ***a single score alone*** (e.g., 80% of patients improved in function) without additional information on case mix ***cannot reveal potential disparities in outcomes across population subgroups, regardless of whether the score is only clinically adjusted*** (the current practice) ***or adjusted for both clinical and SDS factors*** (when appropriate).

As recommended by the Expert Panel and in prior NQF projects, identifying disparities requires additional information and analysis by the relevant sociodemographic factors (e.g., stratification) if the question is: *how do the outcomes of patients with different characteristics compare (either within units or in the population across units)*? Therefore, risk adjustment that includes sociodemographic factors does not change the fact that additional methods (e.g., stratification) as recommended are needed to identify disparities. However, a by-product of adjusting for sociodemographic factors is information about whether within-unit disparities exist. *For example, with multivariate modeling approaches, whether or not and to what degree a SDS factor contributes to the variability in outcomes can be determined.*

Does risk adjustment for sociodemographic factors set a lower expectation for outcomes of disadvantaged populations?

With some risk adjustment procedures, observed counts for an outcome are compared to "expected" counts, which are based on the average experience for patients with similar characteristics. When the average outcome rate for patients with certain characteristics is worse than that for other patients and this value is used to adjust performance scores, some are concerned that it sets a lower standard for the group of patients who experience worse outcomes.

In probability theory, the term "expectation" has a specific technical meaning that differs from its usage in everyday discourse. Generally speaking, it is the value of a random variable that would be observed on average in a large series of repeated trials or random samples. In the context of indirect standardization, the term "expected rate" has a similar technical meaning. It may be loosely translated as describing the "average" or "typical" outcomes for a given case mix. Importantly, the term "expected" is not intended to convey a judgment that "average" or "typical" outcomes are morally acceptable.

Although statisticians use the term "expectation" in this narrow technical sense, it is important to ask what (implicit or explicit) value judgments are reflected in the various accountability initiatives that make use of risk-adjusted performance measures. Policy concerns about accepting or institutionalizing the status quo would not necessarily impact the choice of statistical methodology for risk adjustment and performance measurement, but might reasonably impact decisions about the design of accountability initiatives and the allocation of pay-for-performance (and other) incentives. Nevertheless, risk adjustment does change performance scores (depending on the mix of patients) because the intent is to answer the question: ***how would the performance of various units compare if hypothetically they had the same mix of patients*****?** If the interest is in the question: ***how do the outcomes of patients with different characteristics compare (either within an individual unit or at the population level)***, then performance data stratified by the relevant factors are needed.

Does risk adjustment for sociodemographic factors reduce the incentive to improve care for disadvantaged patients?

There is an expressed concern that socicodemographic adjustment will raise the performance status of some units with a large share of disadvantaged patients from "substandard" to "average" or "average" to "good" and this will lessen the incentive to improve care at those units. There is a parallel concern, sometimes expressed as "masking disparities", that the poor outcomes of such units will be labeled as average – the same label as for a unit producing better outcomes but for a less disadvantaged group. It is unknown whether such a change in labeling will have an impact on motivation to improve, but there is, of course, still an opportunity for such a unit to raise itself to a "superior" level by implementing solutions to problems that affect outcomes for its disadvantaged patients. Motivation to improve is also influenced by the structure of formal (e.g., financial) and informal (e.g., reputational) incentives. We do not know which of these incentives is more motivating.

In general, when there are different categories of patients, the largest group will have the greatest impact on a performance score. Therefore, any improvement in performance in the largest group will improve the overall performance score by a greater amount than a similar improvement in the smaller group. Units interested in improving their overall performance score likely will focus on improvements

affecting the largest number of their overall population of patients. Sociodemographic risk adjustment (or not) does not change this tendency.

If a performance measure is SDS-adjusted, it means that there is a difference in outcomes for one or more specific sociodemographic factors. This signals a need to review data for sociodemographic subgroups to identify opportunities for improvement. The requirement that specifications must also include instructions for stratification of a clinically-adjusted version of the measure based on the sociodemographic factors used in risk adjustment will allow identification of, and facilitate reduction of, sociodemographic disparities. If a unit's case mix includes a high proportion of disadvantaged patients, it will need to address the special needs of that population in order to improve its overall performance score.

When measures that are adjusted for SDS are implemented, the risk adjustment coefficients should be updated on a periodic basis. Thus, improvements in equity of outcomes or processes (that is, reductions in average within-unit quality differences) will be reflected in updated model coefficients and the effects of adjustment would diminish.

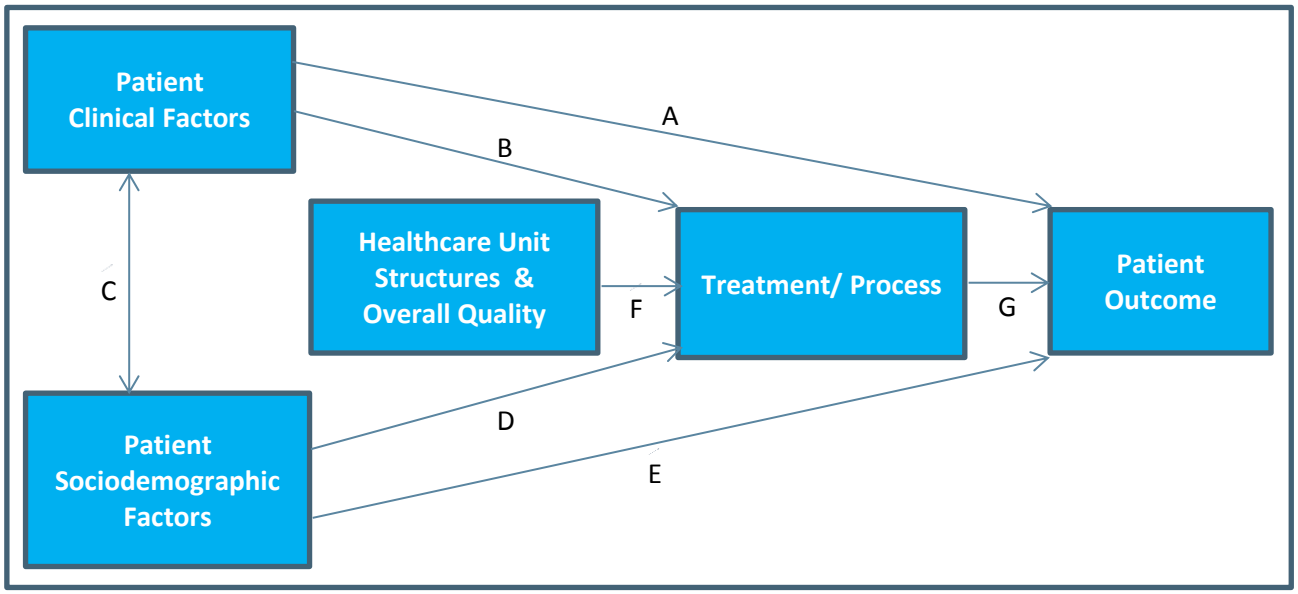
Finally, by appropriately risk-adjusting performance measures, units that have a high proportion of disadvantaged patients and are achieving better outcomes with those patients will be identified as examples for what can be achieved and a source of information about best practices. Their better outcomes might otherwise be masked by the absence of sociodemographic adjustment. Likewise, units achieving good outcomes, but with a low proportion of disadvantaged patients, are less likely to be identified as the best performers.

Does risk adjustment for sociodemographic factors mask disparities in quality if the reason sociodemographic factors affect an outcome is because of the care received?

There may be multiple and complex relationships between sociodemographic factors and outcomes. Following is a simplified path diagram for the effects of patient factors on outcomes.

The objective of performance measurement is to assess overall unit quality through its effects on measurable treatments and processes, as well as its effects on outcomes (path F to G). However, the inference about quality may be confounded by clinical characteristics of patients that affect success in implementing treatments/processes (path B) and/or directly affect outcomes, or through mechanisms not involving the healthcare unit (path A). ***Exactly the same causal relationships hold for sociodemographic characteristics*** of patients (paths D and E), if they affect treatment/process or outcomes.

Figure 1. Causal Paths



Not only should treatment variables be excluded from risk adjustment, variables in the causal pathway between treatment and outcome (path G) should also be excluded because they can distort differences in outcomes by "adjusting away" the treatment effect of interest. For example, one would not adjust for a complication that arises after treatment begins. Even if a variable that occurs after beginning care with the unit does not directly cause outcomes, adjusting for it may cause bias (i.e., systematic deviation from the true value) if the variable is caused by the treatment and is correlated with the outcome. ***These concerns about variables that occur after treatment begins do not apply to sociodemographic factors that are present prior to treatment*** because such factors logically cannot be affected by the healthcare unit – i.e., the healthcare unit cannot affect the patient's level of income or education.

Healthcare unit structures reflect their capacity deliver quality treatments and processes (e.g., staffing numbers and expertise, financial health, performance on other quality measures). External factors can also affect the healthcare unit's capacity to deliver quality care (e.g., area pool of healthcare workers, public funding). These are unit characteristics, not individual patient characteristics used in risk adjustment procedures to account for differences in patient case mix across units.

In general, the path or mechanism of action for a patient factor's effect on an outcome does not need to be known in order to consider it a potential confounder to be assessed for risk adjustment. Adjustment for a variable might make sense if it is a direct cause, an indirect cause, or serves as a surrogate for a cause for which data are lacking. Inferences about comparative quality of healthcare units can be made only IF the potential confounding effects of the relevant factors are controlled (i.e., adjusted), regardless of the path or mechanism.

However, the concern remains that the reason for poorer outcomes on average for patients with particular SDS factors is that disadvantaged patients systematically receive poorer quality care than other patients through either:

- all or most healthcare units provide worse quality care to disadvantaged patients compared to other patients within the same units; or
- all or most disadvantaged patients primarily receive care from poorer quality healthcare units.

Does risk adjustment for sociodemographic factors set a different standard if disadvantaged patients are concentrated in lower quality units?

The above concern about accounting for sociodemographic factors in a risk adjustment procedure can be further elucidated by distinguishing two sources of variation in outcomes across subgroups. Disparities in outcomes for disadvantaged patients can be caused by differences within and between units. Disparities in outcomes are a combination of two components:

- "disparities within": members of disadvantaged groups have worse outcomes than other patients within the same unit (could be due to a variety of reasons);
- "disparities between": members of disadvantaged groups receive care from units where a group of patients would experience inferior care (measured by other processes or outcomes) compared to a group of patients with similar clinical and sociodemographic characteristics receiving care at other units (some refer to this as a contextual effect).

Either or both of these mechanisms can be at work in any dimension of quality, and their relative importance varies across measures and population subgroups. However, ***only the within-unit effects are adjusted for in a risk adjustment procedure because these are the ones that are related specifically to patient characteristics rather than differences across units***.

Adjustment for sociodemographic factors will not mask disparities in quality of care, provided that the risk-adjustment variables are measured at the patient level. The effect of those variables on individual patient outcomes then can be estimated, as long as there is variation in patient characteristics within units. While adjusting for sociodemographic variables will result in a different probability of an outcome it just reflects the patient-based risk reality like any other patient comorbidity. The end result is that each unit's performance score will be compensated for the estimated effects of the sociodemographic factors in proportion to the number of patients in the sociodemographic categories, where those estimated effects are based on the experience of all units in the model. This is appropriate in the context of comparative performance assessment when addressing the question: ***how would the performance of various units compare if hypothetically they had the same mix of patients*****?** An illustration of the effect of risk adjustment appears in Section 5.

While patients with certain characteristics may tend to concentrate differently across units (this establishes a fundamental requirement for risk adjustment), generally there is enough overlap of factor types across units, so that a model correctly estimates the necessary compensation for the disproportionate concentrations. If units differ in quality after adjusting for globally-estimated factor effects, this will be reflected in the profiling results of performance scores.

On the other hand, if members of disadvantaged subgroups tend to be concentrated within units that are overall of lower quality, then methods that ignore such systematic between-unit (contextual) differences can produce biased (i.e., systematic deviation from the true value) unit comparisons. For example, patients without insurance may have poorer outcomes, but this may be in large part because units that treat large numbers of uninsured patients have correspondingly fewer resources, leading to lower quality care for all patients treated by the unit, not just those uninsured. If concentration in low-quality units is a concern, methods exist to appropriately evaluate and address this source of

confounding. The possibility of such clustering of disadvantaged subgroups within lower quality units should be addressed by developers. Examples of methods that can address these issues are given below and in Appendix E .

- The between-unit differences can be controlled for or analyzed using various statistical methods such as including dummy variables for each unit or a unit-level variable that represents the same factor (e.g., percentage of low income patients). If a unit-level factor has an effect that is substantial relative to the patient-level effect, including only a patient-level covariate may result in adjustment for differences in quality of treatment.

  In theory, a patient-level factor could have a strong association with an outcome when between-unit effects are excluded from the model, but a negligible association after adding unit variables for each unit (contextual variables). This would occur, for example, if care for poor and nonpoor patients is similar within each unit but the poor receive care at lower-quality units. In that case, the sociodemographic factor is not a confounder when comparing outcomes across units and efforts to adjust for this factor when comparing outcomes across units may not be needed.

  It is important to distinguish *controlling* for unit effects when estimating within-unit (individual-level) effects, as discussed in the preceding paragraph, from *adjusting* for effects of unit characteristics when reporting quality. The latter is not the intent of risk adjustment, the goal of which is to control for confounders in order to identify the treatment effect of the unit. When unit-level variables are used, they must be used appropriately so as not to adjust performance scores for between-unit differences in quality, which is what you are trying to identify. It is beyond the scope of this paper to provide detailed guidance on statistical methods; however, if unit-level variables are included as described here, procedures for computing the estimated performance score would be different than when only patient-level characteristics are used.

- In the illustration of indirect standardization in Section 5 of this report, if there is concern about concentration of low-income patients in low-quality units, direct standardization would produce valid estimates of healthcare unit performance despite low-income patients being concentrated among units of lower overall quality (see Appendix F). However, if concentration in low-quality units is not a concern, indirect standardization would produce valid estimates of performance. The key point is that just as estimates of unit effects should be controlled for possible confounding by patient characteristics, estimates of the direct effect of patient characteristics (observed within unit) should be controlled for unit effects.

These methods are only mentioned as an indication of the kinds of methods that are relevant and are not a replacement for more technical discussions of various methods. The analyses of within- vs. between-unit effects can be reported and discussed in the measure submission so that reviewers understand these relationships for the specific performance measure. This is an example, where the analyses for SDS adjustment could potentially be used to reduce disparities by identifying the ways in which SDS affects outcomes.

Limitations of Risk Adjustment

Risk adjustment procedures are not perfect even with attention to rigorous methods and principles. Risk-adjusted scores may give a false sense of security and the details warrant close review. The following limitations are acknowledged.

- Data for a potential SDS risk factor with a strong conceptual relationship to the outcome or process being measured, even when based on prior research, may not be available for adjustment. This is not unique to SDS and also occurs with some clinical factors (e.g., stroke severity).
- Even if analyses can identify that an SDS factor exerts its effect on outcomes primarily due to differences in unit quality, the reasons for the differences cannot be determined without additional study. For example, it would not be known whether the presumed quality differences were due to direct action or inaction on the part of the healthcare teams or influenced by lack of public support of safety net providers and insufficient resources to address the increased complexity of disadvantaged patients or to recruit healthcare workers, etc. These unit characteristics or factors would not be included in risk adjustment procedures to account for differences in patient case mix as discussed above. However, community-level factors could be critical for policy considerations.
- A fundamental requirement for risk adjustment is variation in the prevalence of a factor across units; however, if a particular unit has a very different mix of patients from the average mix, the risk adjustment procedure may not adequately adjust the performance score.

Conclusion and Implications

- With appropriate selection of risk factors and risk adjustment methods, sociodemographic-adjusted scores do not mask disparities or differences in quality.

- Based on epidemiologic principles related to confounding and statistical theory of causal inference, the specific path or mechanism for the effect of a SDS factor does not need to be known. However, the requirement for a conceptual and empirical relationship to the outcome (or process) of interest, as well as the other guidelines for selecting risk factors discussed in the next section, will determine whether a sociodemographic factor should be included.

- When considering sociodemographic adjustment, the concern of disadvantaged patents being concentrated in overall lower quality units can and should be empirically tested and if necessary, addressed in the method used for adjustment.

- Risk adjustment does change performance scores if the proportion of patients with various characteristics is different from the average. This is appropriate if the intent is to answer the question: ***how would the performance of various units compare if hypothetically they had the same mix of patients*****?** Regardless of whether the risk adjustment procedure only includes clinical factors or includes both clinical and sociodemographic factors, an adjusted score is <u>not</u> designed to answer the question ***how do the outcomes of patients with different characteristics compare (either within an individual unit or at the population level)*****?** If the interest is in the second question, then data stratified by the relevant factors are needed.

The recommendation regarding sociodemographic adjustment includes the requirement for a conceptual and empirical relationship to the outcome (or process) being measured. Conceptual

considerations may include whether the effect of SDS is primarily mediated through quality of care and questions such as whether there is any reason to think that a central line infection acquired during a hospitalization is influenced by race or income. This is discussed in Section 6.

Although it may be possible to provide some rare but real or simulated examples illustrating some level of presumed failure (to prevent incorrect inferences about quality), that would be the exception rather than the statistical rule. It is not possible to create rules that would accommodate all possible scenarios regarding the use of sociodemographic risk factors. The guidelines for selecting risk factors, beginning with a conceptual and empirical basis, along with statistical and epidemiological theory and practices, provide a sound basis for making those determinations.

Section 5: Effect of Risk Adjustment

Risk adjustment refers to statistical methods to control or account for patient-related factors when computing performance measure scores, including methods such as multivariable models, indirect standardization, or direct standardization. The methodological basis for risk adjustment is presented in Section 4. The result of the statistical procedure is an adjusted overall performance score that takes into account the presence of patient-related factors. Generally, healthcare units serving higher-than-average-risk patients will have adjusted scores that look better than their raw scores; the reverse will be true for units serving lower-than-average-risk patients.

An important goal of risk adjustment is to "level the playing field" when making conclusions about quality of care or performance. That is, the performance scores should not simply be due to differences in the severity or complexity of the patients served. As noted above, the guidelines for selecting clinical and health status risk factors apply to sociodemographic factors. Therefore, without controlling for sociodemographic factors that have a conceptual and empirical relationship to the outcome or process, the inference from the performance score would be incorrect in the context of comparative performance assessment where the central question is: ***how would the performance of various units compare if hypothetically they had the same mix of patients*****?** Sociodemographic factors can contribute to the severity and complexity of the patient population served. Healthcare units with a disproportionate share of disadvantaged patients will appear to provide lower quality care than they actually do, and vice-versa simply as a function of their case mix.

The following illustration is based on one approach to adjustment — indirect standardization. (See another illustration for direct standardization in Appendix F.) With indirect standardization, an expected number of outcomes is determined by applying stratum-specific rates determined from all patients in the reference population to a unit's number of cases in each stratum.[5] An observed-to-expected ratio is then used to compute a standardized or risk-adjusted rate. Multivariable statistical models are an extension of indirect standardization based on the same concepts.

The table that follows illustrates risk adjustment using indirect standardization. This hypothetical illustration does not use actual data and is simplified with just two levels for a sociodemographic factor and numbers chosen for easy computation. ***For purposes of this illustration, one should assume that the sociodemographic risk factor meets the guidelines for selecting risk factors presented in Section 6 and accepted principles regarding confounding discussed in Section 4***. The key points are illustrated in the top of the table — rows 1-6; details about the calculations are provided in rows 7-10.

- The initial scores (row 3) are already adjusted for clinical factors. We will call the performance measure "mortality rate", but it could represent any relatively rare adverse event.
- In this hypothetical example, the national mix of patients is 80% average-high income and 20% low income. The national average experience for mortality is 2% for average-high income patients vs. 3% for low-income patients. Assume that this rate is already adjusted for relevant clinical factors.
- Comparing the overall computed mortality rates that are only clinically adjusted (row 3 labeled "All Patients", unit A has the lowest rate, followed by units B and C (2.2%, 2.6%, and 2.9% respectively). This is an example of the current situation for performance measures, in which clinical adjustment is done, but SDS adjustment is not done.

Table 2. Illustration of Risk Adjustment Using Indirect Standardization

1		All Patients in National Population		Unit A		Unit B		Unit C	
2	**SDS Strata**	Patient mix N/ Percent	**Clinically adjusted deaths** n/Percent	Patient mix n/ Percent	**Clinically adjusted deaths** n/Percent	Patient mix n/ Percent	**Clinically adjusted deaths** n/Percent	Patient mix n/ Percent	**Clinically adjusted deaths** n/ Percent
3	**All patients**	1,000,000 100%	22,000 **2.20%**	1000 100%	22 **2.20%**	1000 100%	26 **2.60%**	1000 100%	29 **2.90%**
4	**Average to High Income**	800,000 80%	16,000 2%	800 80%	16 2%	400 40%	8 2%	400 40%	8 2%
5	**Low Income**	200,000 20%	6,000 3%	200 20%	6 3%	600 60%	18 3%	600 60%	21 3.5%
6	**Income-adjusted rate**				**2.20%**		**2.20%**		**2.45%**
7	**Calculation Details**								
8	Expected deaths Sum of: National stratum rate * unit number of patients in each category				2%*800 + 3%*200 = **22**		2%*400 + 3%*600 = **26**		2%*400 + 3%*600 = **26**
9	Standard ratio = clinically adjusted/ expected deaths				22/22 = **1.0**		26/26 = **1.0**		29/21 = **1.115**
10	**Income-adjusted rate** Ratio *National rate				1.0* 2.2% = **2.20%**		1.0* 2.2% = **2.20%**		1.115* 2.2% = **2.45%**

- Unit A's sociodemographic case mix is the same as the national mix. Its performance is also the same as the national average for both the average-high and low-income categories (2% and 3% respectively). In both case mix and performance, then, it is exactly average. Adjustment for income using this method (result in bottom row) does not change its rate (2.20%).

- Unit B's sociodemographic case mix (rows 4-5) has a higher proportion of low-income patients, but its performance is exactly the same as the national average as well as that of unit A for the two income categories (2% and 3% respectively). With its performance score not adjusted for income (2.6%), its performance appears to be "worse" than unit A, but in fact it is not. In its rate adjusted for clinical factors and income (row 6), its performance score is identical to that of Provider B (2.2%). This reflects that the question being addressed is: *how would the performance of various units compare if hypothetically they had the same mix of patients*?
- Unit C has the same sociodemographic mix of patients (rows 4-5) as unit B, but its performance is worse for the low-income group (3.5% vs. 3%). Its income-adjusted rate (row 6) is higher than unit B's income-adjusted rate, reflecting its poorer performance for its low income patients.

There are three important points to emphasize about this example.

- First, adjustment for income in this particular illustration does not "adjust away" the differences in results achieved between unit B and unit C. Unit C still has a worse performance score than either A or B after adjustment.
- Second, income disparities are clearly visible in the data for each stratum (rows 4-5), and they are actually a key part of the middle steps of the indirect standardization calculations. This is the data that would be available to identify disparities both across and within units with the recommended stratification.
- Finally, all three units in this scenario may have incentives to improve. In a "star system" of rankings, Units A and B might have "three-star" designation because their performance is just average. If rewards are given for four- or five-star performance, they both need to improve. Unit C may have a two-star designation depending on how cut points are set, but it also has a clear incentive to improve. It may be the case that both Unit B and Unit C find that their best opportunity for overall improvement is to improve care for their low-income patients because they comprise a substantial proportion of their population.

Neither the observed nor adjusted performance rates alone can provide any information on disparities. Without the specific information on performance for income subgroups, the overall performance rates neither identify nor mask disparities. The subgroup scores that are included in this method do reveal the disparities, though. This particular adjustment method meets the Panel's general principles of transparency, attention to disparities, and validity and fairness of performance assessment.

Risk adjustment is not perfect and the same limitations of risk adjusting for clinical factors applies to sociodemographic factors — that is, when patient mix affects outcomes and differs widely across units, risk adjustment may not completely adjust for those differences.[5] For example, if the national mix of patients is 20% low-income, but a particular unit's patient mix consists of 90% low-income patients, a national adjustment model may not be able to completely account for such a large difference in case mix. Therefore, risk adjustment does not necessarily preclude using additional methods when comparing performance such as constructing peer groups for comparison as described below.

Alternatives to Risk Adjustment

Stratification

Stratification refers to computing performance scores separately for different strata or groupings of patients based on some characteristics(s) — i.e., each healthcare unit has multiple performance scores (one for each stratum) rather than one overall performance score. For example, strata could be constructed based on poverty level and performance scores computed for each stratum. Sometimes stratification is considered a type of risk adjustment as a means to making like comparisons; however, the Expert Panel thought statistical procedures such as multivariable models and stratification were so different, that they are considered separately. With stratifications, performance is reported and can be compared for subgroups of patients with similar levels of risk or sociodemographic characteristics. It offers two advantages. Stratification by sociodemographic factors allows identification of disparities in healthcare for certain subgroups of patients because scores are associated with the particular factor. In essence, stratification "unmasks" healthcare disparities by examining performance for groups who have been historically disadvantaged compared to groups who have not been disadvantaged.

An illustration of stratification appears in the table that follows. Note that stratification is essentially the first step in adjustment as illustrated in the example above. Stratification is most likely to be useful when examining performance for groups where substantive differences in performance have been observed. It is particularly useful for providing finer-grained information and most notably for assessing and addressing disparities.

The biggest barrier for using stratification alone for accountability applications is one of feasibility. Each healthcare unit's patient population is divided into the specified categories, thus reducing sample sizes available for analysis in each category. Sample size affects reliability and the ability to distinguish differences and make accurate inferences. If there is more than one relevant sociodemographic factor (e.g., race, ethnicity, income, language, etc.) then stratification becomes much more complex, increasing the number of categories and further reducing sample size in each "cell" of the resulting matrix of stratification factors and levels. Combining individual factors into composites may address this problem to some degree, but stratification by itself does not address the problem of needing a single performance score for each unit for a given measure in order to use in either public reporting or pay-for-performance.

Table 3 shows a very simple example of stratification. A single sociodemographic variable (income, for example) is divided into three levels, and patients are assigned to one of the three levels. The size of the population at the national level may be in the hundreds of thousands or millions for each of the three strata, so performance rates are very reliable. In this illustration a higher performance rate is desired and indicates better quality. Unit A has reasonably large sample sizes in each stratum, and performance scores close to the national average in each. Its performance, for each stratum, would probably be identified as average. There is no direct way, in this example, to make a judgment about unit A's overall performance, although its higher proportion of patients in the "low" stratum would tend to make its overall performance appear to be worse than average, as shown earlier in Table 2. Disparities in performance across the three strata are evident, and are essentially the same as the disparities found at the national level. Unit B has a much smaller sample in each stratum, and also lower performance scores in each. It actually has lower disparities across strata than unit A, but its overall performance score would be worse. (Note, though, that the stratified report does not actually provide an overall score.) However, the small sample sizes in each cell may make it difficult to identify the performance as significantly worse than either unit A or the national average.

Table 3. Illustration of Stratification

	National		Unit A		Unit B	
Sociodemographic Stratum	Percent of patients	Observed rate	n/ percent	Clinically-adjusted rate	n/ percent	Clinically-adjusted rate
Low	30%	63%	500 50%	65%	20 20%	60%
Moderate	50%	72%	400 40%	70%	50 50%	65%
High	20%	85%	100 10%	83%	30 30%	67%

Peer Groups for Comparison

Peer groups for comparison refers to creating peer groups of healthcare units caring for a similar mix of patients, within which to examine performance scores. It could facilitate comparisons of units with similar resources, e.g., VA sites with VA sites, or federally qualified health centers with each other. Depending on how the peer groups are constructed, it also tends to match patient populations, e.g. proportion of uninsured patients or those covered through Medicaid. In this approach, performance scores for individual units are neither adjusted nor stratified for sociodemographic factors (using the definition of stratification in this report). Constructing peer groups for comparison occurs after performance scores are computed. This approach avoids the issue of reducing sample sizes seen with stratification. Recently, MedPAC recommended using this approach with the readmission reduction program.[36] Peer groups can help ensure that use of a performance measure to apply rewards or penalties is consistent with program and policy goals. For example, if units caring for a disproportionate share of disadvantaged patients will be disproportionately penalized using unadjusted performance scores, then that may not be a desirable result. Applying the penalty on the basis of performance within groups of "peer units" rather than on the basis of performance relative to the entire universe of units is one way to avoid a disproportionate share of penalties to safety-net units. Adjustment of a performance score for sociodemographic factors would not always or automatically exclude the possibility of using peer groups for comparison.

Table 4 presents a simple example of use of peer groups to establish different reference points for different units, which then could be used to apply financial rewards or penalties or to identify providers as relatively good or bad within that peer group. In this example, units (e.g., hospitals) are grouped into "quintiles" based on the percent of their patients at or below 138% of the federal poverty level. Hospitals in quintile 1 have relatively few such patients; hospitals in quintile 5 have many. The performance measure here is something for which "more is better" — percent of acute stroke patients

arriving at the emergency department within two hours of symptom onset, for example. There are clear disparities in performance on this measure at the national level in this example.

Units A and B are both assigned to a quintile based on their specific percent of patients at or below 138% of the federal poverty level. Unit A is in the middle quintile and unit B is in the fifth quintile. Unit A's performance is a bit worse than the quintile average (62% vs., 65%), so it might be identified as a "below average" performer for its quintile. Unit B's performance is a bit better than average for its quintile (59% vs. 55%), so it might be identified as an "above average" performer for its quintile, even though its performance is worse than A's in absolute terms.

The Panel had generally favorable views of this approach as a method to more fairly apply financial rewards and penalties. However, determining appropriate peer groups can be challenging. The method does not, however, identify disparities in care within units, nor does it indicate whether unit A or B is better than the other if scores had been adjusted for patient income, either for specific subgroups of patients or overall. It is possible, depending on the exact distribution of patients across income strata for the two units, that unit B would have a better score with an adjustment approach like that illustrated in Table 2. Some view this approach as more explicitly setting different benchmarks for healthcare units based on the proportion of disadvantaged patients served.

Table 4. Illustration of Peer Groups for Comparison

Quintile based on percentage of low income patients	Number of Units	Quintile cut point of percentage of patients at or below 138% of poverty	National Average clinically adjusted rate for units in quintile	Unit A Clinically adjusted rate	Unit B Clinically adjusted rate
National	3000		70%		
1st Quintile	600	10%	75%		
2nd Quintile	600	20%	70%		
3rd Quintile	600	30%	65%	62%	
4th Quintile	600	40%	60%		
5th Quintile	600	55%	55%		59%

The three general approaches described here — statistical risk adjustment, stratification, and peer groups for comparison — are not mutually exclusive. They could be used in combination or in all three ways for a given performance measure, with the specific analytic approach chosen for a specific analytic or program purpose. In an analysis focusing on the presence of sociodemographic disparities in care, for example, stratification would be the natural first-choice approach, as it provides the clearest and simplest information about performance in relation to a particular sociodemographic factor. For some program purposes, like application of a hospital readmission penalty, a peer-grouping approach might be simplest and most desirable. Each has different strengths and limitations. The Panel concluded that different approaches serve different purposes. A strong majority of Panel members did not think, however, that either stratification or creating peer groups would be adequate for all "accountability" measurement purposes. When single performance scores are interpreted as indications of underlying quality of care, the large majority of the Panel thought that statistical adjustment for relevant sociodemographic factors when indicated would be necessary to support valid inferences about quality and that stratification was needed to assess and address disparities.

Section 6: Guidelines for Selecting Risk Factors

The Expert Panel reviewed the accepted guidelines for selecting clinical or health status risk factors and their rationales. The Panel determined that these same guidelines may also be applied to sociodemographic factors. As indicated in Recommendation 1, several conditions must be met before a performance measure is adjusted for SDS. These conditions are consistent with selecting clinical risk factors. Each performance measure must be assessed individually.

Conceptual and Empirical Relationship for SDS Adjustment

The first condition for selecting risk factors is that a conceptual relationship and an empirical relationship exists between the specific risk factor and the outcome (or process) being measured. A conceptual relationship refers to a logical theory or rationale that explains the association. The conceptual basis may be informed by prior research and/or healthcare experience related to the outcome of interest, but does not require a direct causal relationship (i.e., it could be a direct cause, an indirect cause, or serves as a surrogate for a cause for which data are lacking). An empirical relationship means that there is a statistical association between variables for the risk factors and the variables for the outcome.

Not all outcomes or processes of care are affected by sociodemographic factors. For example, outcomes and processes such as the outcome of central line infection occurring during a hospital stay, or the process of administering the correct medication at the correct time during a procedure, would not have a conceptual reason for a relationship with sociodemographic factors. One would expect the same things to be done, and the same results obtained, for any and all sociodemographic subgroups. Further, not all sociodemographic factors may affect all outcomes. For example, improvement in ambulation has no conceptual relationship to race, but does to age.

The recommendation on SDS adjustment also allows that SDS adjustment might be unnecessary or inappropriate based on conceptual reasons or empirical evidence. Some examples include whether the influence of the SDS factor is primarily through quality of care delivered, empirical analyses that indicate the potential factor does not account for variation in the outcome being measured, or empirical analyses that indicate the effect is through disadvantaged patients being clustered in poorer quality units (as discussed in Section 4).

An assessment of a conceptual relationship between a 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. That is, situations in which the SDS factor leads to the delivery of inferior care processes, which in turn affect the outcome. An obvious example is unequal treatment to patients with a particular characteristic such as race or homelessness where they are consistently skipped in routine screening for hypertension, which leads to higher rates of blood pressure greater than 140/90. If this was the general and pervasive practice for those patients, it could be reason enough to not consider those factors for risk adjustment, even if they have an empirical association with the outcome. The mechanisms for the effect of specific SDS factors on specific outcomes may be complex, involving multiple paths, and essentially unknown without additional study. As discussed in the methodological basis for risk adjustment, the exact mechanism of the effect on an outcome does not need to be known in order to consider it or reject it for risk adjustment, and adherence to the epidemiological and statistical methods and principles related to confounding and the guidelines for selecting risk factors will inform whether it is included or excluded.

Some potential questions for identifying a conceptual basis for adjusting a performance measure for sociodemographic factors include:

- Does prior research indicate a relationship between SDS and the outcome?
- Is there a logical relationship or theory about the relationship between SDS and the outcome?
- Is there a significant passage of time between the healthcare unit intervention and measured outcome in which other factors may have an effect?
- Do patient actions or decisions influence the outcome or process and are the decisions affected by SDS (e.g., ability to purchase medications)?
- Does the patient community have an influence (e.g., distance to pharmacies, groceries, healthcare services)?

If a conceptual relationship exists between a patient-level sociodemographic factor and outcome, it should be tested empirically if data are available. The Panel did not specify, and does not recommend, any particular analytic approach with which to assess empirical associations between sociodemographic factors and outcomes, nor any specific cutoff or threshold value to use for declaring the presence of an association. A common method to identify an empirical relationship is to assess the correlation between the two variables. For example, as income increases, mortality decreases. If the basic conditions for conceptual and empirical relationship are met, then SDS factors will be assessed for inclusion in risk adjustment procedures following the remaining guidelines for selecting risk factors.

Some have advocated that sociodemographic factors affect clinical and health status and therefore, may already be accounted for through those risk factors. That is a possibility that can be tested empirically. It is also important to consider that if sociodemographic factors lead to less use of healthcare services, data on health status and clinical conditions prior to the start of care may not exist to the same degree for disadvantaged patients as for those who use healthcare services more frequently.

As indicated in Table 5. Risk factors do not need to meet every guideline and developing a risk model is an iterative process that at times requires weighing various trade-offs.

Table 5. Guidelines for Selecting Risk Factors for Adjustment

Guideline	Rationale	Clinical/ Health Status Factors[b]	SDS Factors[c]
Clinical/conceptual relationship with the outcome of interest	Begin with conceptual model informed by research and experience	✓	✓
Empirical association with the outcome of interest	To confirm conceptual relationship	✓	✓
Variation in prevalence of the	If there is no variation in prevalence across	✓	✓

[b] Examples of clinical and health status factors include comorbidity; severity of illness; patient-reported health status, etc.
[c] Examples of sociodemographic factors include income; education; English language proficiency, etc.

Guideline	Rationale	Clinical/ Health Status Factors[b]	SDS Factors[c]
factor across the measured entities	healthcare units being measured, it will not bias performance results		
Not confounded with quality of care – risk factors should:	Trying to isolate effects of quality of care	✓	✓
• be present at the start of care and	Ensures not a result of care provided	✓	✓
• not an indicator or characteristic of care provided (e.g., treatments, interventions, expertise of staff)	Although these could explain variation in outcome, in performance measurement the goal is to isolate differences in performance due to differences in 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	✓	✓
Potentially, improvement of the risk model (e.g., risk model metrics of discrimination – i.e., sensitivity/specificity, 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 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	✓	✓

Information Submitted for Review and Evaluation for Potential Endorsement

The Expert Panel recognized that developing adjustment strategies for performance measures is an iterative process involving a conceptual basis and empirical analyses resulting in multiple decisions to arrive at a final risk adjustment procedure. There is more than one appropriate way to accomplish adjustment. Therefore, NQF should not be prescriptive regarding methods for adjustment or specific SDS variables. However, steering committees and stakeholders need to have sufficient information to evaluate performance measures for endorsement. When a measure is submitted to NQF for potential endorsement, it is important that the developer's rationale regarding adjustment for sociodemographic factors be transparent and open to review and evaluation.

In addition to the adjustment methods, factors, and rationale, the developer should discuss the potential risk of misuse of the measure. NQF already requires information on current and planned use of measures. The developer has detailed knowledge about the limitations of the performance measure that could impact its use in accountability applications.

The Expert Panel identified the following as important aspects for reviewers to evaluate whether SDS adjustment is appropriate.

- Conceptual description (logical rationale or theory informed by literature and content experts) of the causal pathway between sociodemographic factors, clinical factors , quality of care, and outcome

- Sociodemographic variables that were available and analyzed. For example:
    - Patient-level factors
        - Patient-reported (e.g., income, education, language)
        - Proxy variables when sociodemographic data are not collected for each patient (e.g., based on patient address, use of census tract data to assign to a category of income, education, etc.), conceptual rationale, and analysis
    - Patient community factors assigned to patients from the specific community (e.g., crime rate, percent vacant housing, smoking rate, level of uninsurance) — see discussion of community-level factors below);

- Analyses and interpretation resulting in decision to include or not include SDS factors. For example:
    - Prevalence of the factor across measured entities
    - Empirical association with the outcome
    - Contribution of unique variation in the outcome
    - Assessment of between unit effects vs. within unit effects as discussed in the methodological discussion in Section 4

- Current and planned use of the measure and a discussion of risks for misuse of the specified performance measure

Section 7: Specific Sociodemographic Factors to Consider for Adjustment

Adjustment of the performance score generally involves patient-level data for the risk factors — i.e., an individual patient's diagnosis, lab value, income, education, etc. Although the Expert Panel agreed that performance measures should be adjusted for sociodemographic factors when appropriate, it also recognized the data challenges that constrain adjustment. Data about patient sociodemographic factors other than age and sex often are not collected, or not standardized sufficiently for use in performance measurement.[37] Collection of race and language by healthcare units is growing but SES-related data are not widely collected. Therefore, data availability is a critical consideration. Besides overcoming prior assumptions, data constraints may be the biggest barrier to adjustment for sociodemographic factors and will require further initiatives to define standards and to implement data collection.

When sociodemographic data are not collected for each patient, other methods may be used to assign a value for each patient (e.g., based on census data for the patient's home address or Zip Code). Just as whether sociodemographic variables are used in adjustment should be based on conceptual relationships, use of proxies for patient sociodemographic data should also have a conceptual basis. For example, area-based data could be assigned as a crude proxy for individual SES, or as Krieger[38] suggests, could characterize the patient's environment. That is, if one uses census data on income for a given patient's neighborhood, one can either be saying "I think you're probably poor because you live in this neighborhood" or "You live in a neighborhood with mainly poor people in it".

The Expert Panel identified potential sociodemographic factors that might be useful for adjustment and discussed some of the pros and cons when considering those factors for adjustment. However, the Panel did not recommend specific variables to be used — that will depend on applying the guidelines for selecting risk factors for a particular performance measure, as well as on data availability.

Age is a clinical variable as well as a demographic variable. Physiologic changes accompany age and the probability of disease increase with age. It is already included in many clinical risk adjustment procedures and should continue to be utilized as the conceptual and empirical relationships with the measured outcome indicate.

Socioeconomic Status (SES)

SES arguably represents a fundamental determinant of health,[39] and access to and use of health care.[40] SES represents a multidimensional construct that has been traditionally measured based on income, education, and occupation (although much greater attention has been given to the first two dimensions).[41]

Income

Income is a key dimension of SES. It affects health over the life course and healthcare access and affordability. These effects have been extensively documented.[10, 42-44] Optimally, household income should be collected directly from patients. This is currently done in selected instances, (e.g. to assess eligibility for charity care, subsidies for health insurance on the exchanges), but it is not widely collected in healthcare. A key barrier is reluctance to asking all patients about their income(potentially resistance from both patients and healthcare units).[45] A second barrier is that income is difficult to measure because household income can come from multiple sources for each person within a household.[46] A full assessment requires multiple questions.

Income variables need to be considered in light of variations in cost of living and purchasing power across the U.S. For national use, consideration should be given to standardization by wage or cost-of-living indexes.

When individual or household income data are lacking, proxies based on residence may be used.[38, 47] Area-level data may be used to either impute characteristics of individuals or to characterize the areas in which people live, and there is a rich literature on pros and cons of either usage. Area-based measures of income can be based on patient addresses geocoded to the Census Tract, Block Group or Block. ZIP codes can be linked to census data; however, ZIP codes are limited because of greater socioeconomic heterogeneity within the area.[38] Smaller, less heterogeneous areas may yield more valid results when used as a proxy for individual income. The Geocoding Project showed findings regarding the association of SES with mortality and with cancer incidence were most consistent when addresses were geocoded at the Census Tract than at the ZIP Code or Census Block Group. Recent developments have improved the matching of addresses to areas and have minimized failures to successfully geocode addresses.[48-50]

Medicaid eligibility or dual eligibility for Medicaid and Medicare often is used as an indicator of low income. Although there is significant heterogeneity in Medicaid eligibility, benefits, and payments between states, it is a verified indicator of low income and the information is widely available. Expanded eligibility for Medicaid through the Affordable Care Act (ACA) represents a verified measure of household poverty (i.e. <138% federal poverty). Currently, 25 states, in addition to the District of Columbia, have opted for expansion. This expansion presumably will grow over time providing for a standard measure of poverty across states with similar eligibility. However, some low-income people will not be eligible for Medicaid with the ACA expansion due to immigration status or other reasons.

Education

Education represents another dimension of SES. It is powerfully related to health, health behavior, and healthcare.[10, 51, 52] Like other measures of SES, it clusters by healthcare organization and provider.[53, 54]

Nearly two decades ago, the National Committee on Vital and Health Statistics proposed that education (i.e., years of schooling) is a core health data element that should be standardized in healthcare and healthcare information fields.[55] Despite this recommendation, education is not widely collected in healthcare outside of patient experience of care surveys (e.g. CAHPS) and is inconsistently collected by clinical personnel as part of the social history of patient that is included in the medical record. In contrast to household income, education may be easier to collect from patients with fewer refusals.[45] Currently, Meaningful Use Stage 3 standards require collection of patient race, ethnicity, and language data and state that electronic health record vendors should encourage and incentivize new types of data collection; however, these standards do not specifically mention patient educational level. Regulations and promotional efforts have fostered collection of race, ethnicity and language among hospitals[56] and health plans.[57] Similar approaches could be used to promote collection of individual patient educational attainment within structured data fields (that can be exported). Until these data become available, area-based measures (discussed in more detail below) may be used as crude proxies.[58, 59] Standardized collection of patient (or parental education) in healthcare would obviate use of imputed measures of patient education. This represents an important priority related to improved measurement of SES in healthcare. An IOM report on optimal social and behavioral measures for collection in EHRs recommended inclusion of patient educational level.

Occupation/Employment

Occupational level represents the third dimension of SES. Employment status is more easily assessed and potentially relevant given its relationship to health insurance, health behavior, and mortality, and represents an additional potential adjustor.[60-63] Existing methods for classification of occupations have limitations.[64] Moreover, relatively little is known about its effect on outcomes independent of other measures of SES and sociodemographic-related factors. Obtaining standardized occupation data from patients generally does not lend itself to single questions.[65]

Sociodemographic Factors Related to SES

Language

Limited English proficiency (including communicating through American Sign Language) contributes to suboptimal healthcare, inadequate informed decisionmaking, poor self-management, and healthcare disparities.[66-72] These barriers persist despite language assistance regulations[73] and the recognition of language differences as barriers to quality and safety by The Joint Commission.[74] A 2009 Institute of Medicine (IOM) report recommended standardized data collection for language in addition to race and ethnicity.[75] Subsequently, progress has been made by hospitals and health plans in the collection of these data using a combination of direct and indirect methods.[76-78]

Insurance

The uninsured disproportionately includes minorities, the poor, those with low education, and those with limited English proficiency.[79] Health insurance is strongly associated with healthcare use, improved preventive and chronic care management, and reduced mortality for children and adults.[10, 80-83] The presence or absence of insurance may be useful for adjusting quality performance measures. An important related measure is under-insurance.[84, 85, 86] Out-of-pocket payments not covered by health insurance affect patients' healthcare decisions, particularly among poorer patients.[85] Optimally, data on the quality of insurance analogous to the designations for insurance purchased on health exchanges, (i.e. bronze, silver, gold, and platinum) could be collected to assess patient under-insurance.

Race and Ethnicity

Race and ethnicity are not and should not be used as proxies for SES; rather, their effects are confounded by SES.[87] That is, income, education, and related factors (including language and insurance) represent key contributors to racial and ethnic disparities in healthcare.[10, 42] Potential mediators of the effect of race on outcomes include source of care,[19, 88] discrimination,[89] and potential differences in biology (including those that are environmentally- or stress-induced). Potential biological effects include high rates of preterm birth among African Americans[90] and differences in levels of glycated hemoglobin between Blacks and Whites.[91-93] For other outcomes, such as hypertension control, there is conflicting evidence as to whether factors such as discrimination, fear of side effects, and/or adherence to treatment plan fully account for disparities in blood pressure control or not.[94, 95] Although some see race/ethnicity technically as no different than other potential confounders, because of the concerns about bias and racism, careful thought, consideration, and a clear rationale should be used when adjusting performance measures for race and/ethnicity. At the same time, reporting of data stratified by race and ethnicity should be encouraged to assess and address disparities in healthcare. Collection of race and ethnicity data is improving, but gaps remain hindering use of these data.[96, 97]

Homelessness

Homelessness is associated with poor healthcare access and high levels of unmet healthcare needs, poor health, and hospital re-admission.[98-101] However, patients frequently are not asked about their housing status, even during hospitalization.[102] Standardized definitions for homelessness have been developed and are used by Housing and Urban Development.[103]

Marital status

Marital status is strongly associated with household income. It is not only related to health behaviors, but to health and mortality, particularly following disruption through divorce or death.[104-109] It is easily and often collected along with other demographic factors in the process of hospital admission or clinic registration. Marital status is also strongly related to the availability of caregiver availability that is known to be related to health outcomes in post-acute settings.

Literacy and health literacy

Literacy (ability to effectively read and write), numeracy (ability to understand and use numbers in daily life),[110] and health literacy (capacity to obtain, process, and understand basic information and services needed to make appropriate decisions regarding health)[111] are associated with educational attainment.[112] Both general literacy (and numeracy) and the related construct of health literacy are strongly associated with healthcare use and outcomes.[113-115] Brief screening tools show promise for health literacy.[116]

Community Variables

Proxy for patient-reported data

Risk factors are considered patient-level characteristics and in that context, "community" refers to the community where the patient resides, not the community where the healthcare unit is located. Community variables could be used at the individual level to characterize the environment in which the patient lives. Community-level variables include the geographic distance to pharmacies, availability of public transportation, types and availability of food outlets, neighbor and social support infrastructure, and availability of parks and recreation areas. These may be as, or more, important than individual SDS characteristics in terms of accounting for access to economic and social infrastructure and health care services, all so important to good health outcomes. In rural communities, this includes the geographic distance to healthcare providers. Other examples include rates of crime or percentage of blue collar or professionals residing in the area.[117, 118] Because multiple variables of social disadvantage by Census area are available, some researchers have used composite measures based on factor analysis.[47, 119] However, a single measure (poverty) may perform as well as composite measures.[38]

Community-based measures of SES have been used to characterize SES of patients in health plans and quantify socioeconomic disparities in quality.[58, 120, 121] The specific variables selected and how they are used should be based on the conceptual model. While community-based measures potentially will misclassify some individuals when used to impute individual-level characteristics due to socioeconomic heterogeneity within the area being measured,[122] they offer the potential for capturing contextual effects beyond individual measures including insurance availability or public support for health care.[123-125]

Community factors affecting the healthcare unit

Some community characteristics are most relevant as characteristics of the healthcare unit, for example, funding for safety net providers (a function of local and state taxing jurisdictions and associated public funding or lack thereof) and the pool of available healthcare workers for employment. Because they are not characteristics of patients, they would not be included in risk adjustment procedures as discussed in Section 4. How these types of factors should be addressed in performance measurement (e.g., in hierarchical modeling approaches) and the implications needs to be further explored.

Potential Mediators of Sociodemographic Factors

There are a number of potential mediators between sociodemographic factors and outcomes. Examples include social support (and its converse, isolation and loneliness),[126, 127] and "patient activation" which refers to patient confidence and skills needed to assume shared responsibility for their health and health care.[128] A range of behavioral factors, including smoking, alcohol use, physical activity, and diet,[129] may be mediators of effects of some sociodemographic factors; however, these are more likely than SDS factors to be included in clinical risk adjustment models, along with self-reported health status.[130]

As previously noted, selection of sociodemographic risk factors should first be guided by the conceptual relationships, but before any analyses of relationships with outcomes can be conducted, the data must be available. There may be several options for operationalizing a sociodemographic concept and the Expert Panel identified some of the pros and cons of various variables to consider when selecting variables for sociodemographic adjustment (see Table 6).

Table 6. Sociodemographic Factors – PRO and CONs

Factors/concepts (specific variables)	**PROs**	**CONs**	**Caveats**
Factors that should be considered, depending on: data availability and the specific outcome or process			
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 all patients • Meaning is not geographically consistent due to difference in costs of living	• For national performance measures, need to consider standardization to account for area wage and cost of living differences
Income in relation to federal poverty level	• Definition is standard • Being used under ACA • Researchers are used to using it	• Doesn't include receipt of other benefits (e.g., food stamps) • Doesn't account for cost of living or community offsets	
Household income	• May be more meaningful than individual income	• Requires assessment of household size	
Medicaid status as	• Relatively easy to collect in	• Eligibility not consistent across	• Potentially

Factors/concepts (specific variables)	PROs	CONs	Caveats
proxy	claims data	states	becomes more useful as more States expand Medicaid to 138% federal poverty level
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 misreporting 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	• Not widely collected by healthcare units • If collected (e.g., in EHR text fields) may not be easily retrievable	
Homelessness	• Strongly associated with health outcomes • Measures something "beyond" income • Current HUD definition	• Multiple other definitions • Data often not collected • Status can change	• Prevalence tends to cluster among safety net healthcare units
Housing instability	• May be better indicator than homelessness which can change	• More difficult to define than homelessness	
English proficiency	• Standard definition exists • Tied to need for translation services/other resource needs and therefore should be collected • Increasingly being collected (required by “Meaningful Use” and some states)		
Insurance status	• Readily available • Some indication of access and resources • Benefit coverage strongly related to affordability	• Wide variability in insurance coverage • Data for under-insurance not widely collected	
Medicaid status	• Readily available	• Not consistent across states	

Factors/concepts (specific variables)	PROs	CONs	Caveats
	• Some indication of limited income and resources		
No insurance	• Readily available • Standard meaning		• Difficult to capture information about these patients (particularly if using claims data)

Community/Neighborhood-level data used as proxy for individual data or as contextual variable	• Many variables available from Census data
	  • Income
	  • Education
	  • Immigration status
	  • Language
	  • Unemployment
	  • Home ownership
	  • Single parents
	  • Others
		• Census data do not include all potentially important variables
		• Residential heterogeneity will affect whether it is a good proxy for data about individuals.
		• 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	• Seen as indicator for other related issues such as poverty, crime, lack of resources		
Contextual- Crime rate	• May be an indicator for other related issues such as poverty, lack of resources		
Other factors that could 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	• Multidimensional construct that typically requires multiple questions • Lack of agreement about how to measure • Not consistently measured	
Living alone	• Available in OASIS data for home health	• Directionality may not be consistent. In some situations such as frailty or impairment, it could be a risk factor. In other situations, it might be an	

Factors/concepts (specific variables)	PROs	CONs	Caveats
		indicator of ability to live alone due to good health and function.	
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 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	• Often collected	• Employment status does not reflect income or availability of insurance • Simple yes/no does not reflect desire/happiness with situation (e.g., retirees may be happy to be unemployed) • Subject to change requiring continuous updating	
Literacy	• This concept may also be able to partially capture health literacy	• No standardized definitions • May be easy to game	If the correlation with education is high, then education could be used.
Health literacy	• Potentially more relevant to healthcare • Three-item and single-item validated questions exist	• Not consistently collected/ available	
Local/state funding for safety net providers (e.g., tax base)	• Affect resources available to safety net providers beyond insurance	• Data not easily collected/ available	• Not a patient characteristic • Risk for unintended consequences (setting a lower standard for poorly supported institutions might

Factors/concepts (specific variables)	PROs	CONs	Caveats
			send the wrong messages to tax payers)
Race/ ethnicity	• Correlated with SES and may be more available than other variables	• May be more correlated with bias	• Should not generally be used as proxy for SES

Section 8: Policy-Related Discussion

Use of Performance Measures in Accountability Applications

NQF-endorsed performance measures are expected to be used in accountability applications such as public reporting and pay-for-performance. The NQF criteria focus on endorsing measures that demonstrate reliability and validity and adequate risk adjustment so that correct conclusions about the quality of care can be made by patients and others. NQF does not set different reliability and validity standards for different accountability applications. As already noted, concerns have been expressed about the policy response to performance results. For example, if providers or health plans serving disadvantaged populations have poorer outcome performance and incur financial penalties, it could worsen disparities in health and healthcare by reducing resources available to care for their patients. Therefore, it is imperative that various accountability applications be assessed for the potential impact on providers and plans caring for disadvantaged populations to identify unintended consequences to patients and to ensure alignment with program and policy goals.

Even if a performance measure is adjusted for sociodemographic factors, it does not rule out the potential need for also creating peer groups for comparisons in various accountability applications. Even when risk adjustment includes relevant factors, it may not fully account for differences in risk when patient mix affects outcomes and differs widely across healthcare units or due to data limitations.

When a measure is submitted to NQF for endorsement, information on current and planned use should be submitted. Currently, NQF criteria and endorsement do not include requirements for or evaluation of procedures for implementation and reporting of the computed performance measure score (e.g., reporting with or without confidence intervals or sample sizes; methods for determining rankings or ratings, statistically significant differences, or incentives and penalties). However, the way a measure is implemented involves multiple decisions that could affect the validity of conclusions (inferences) made about quality of care and create 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 for units above or below a cut point (i.e., confidence intervals for scores above and below a cut point may overlap). Review of the detailed information about the performance measure for potential endorsement provides an opportunity to identify any specific considerations or limitations for use in specific accountability applications.

The Expert Panel recommended that NQF should consider expanding its role to include guidance on implementation of performance measures. Possibilities to explore include:

- guidance for each measure as part of the endorsement process;
- standards for different accountability applications (e.g., use in pay-for-performance versus pay-for-improvement; innovative approaches to quality measurement explicitly designed to reduce disparities).

Some Panel members expressed concern about endorsed measures being used inappropriately, and the Expert Panel recommended that NQF should make explicit the existing policy that endorsement of a performance measure is for a specific context as specified and tested for a specific patient population (e.g., diagnosis, age), data source (e.g., claims, chart abstraction), care setting (e.g., hospital, ambulatory care), and level of analysis (e.g., health plan, facility, individual clinician). This policy is implicit in the

current NQF criteria and process for endorsing a measure as specified and tested, but the Panel expressed concerns about inappropriate application of modifications to endorsed measures.

Use of Performance Measures to Identify and Reduce Disparities

Recommendation 1 acknowledges that when a performance measure is SDS-adjusted, it is disparities sensitive. The second part of the recommendation states: the performance measure specifications must also include specifications for stratification of a clinically-adjusted version of the measure based on the sociodemographic factors used in risk adjustment.

A single performance score (whether adjusted or not adjusted for sociodemographic factors) neither identifies nor masks disparities—that requires the additional information about the characteristics of the patients served. In other words, the current system of performance measurement does not allow disparities to be identified so that they can be eliminated. Doing so requires analysis of performance measures that are stratified as recommended with NQF-endorsed disparity sensitive measures[131]. Hence, the Panel made this recommendation. This approach also helps address concerns about masking performance for disadvantaged groups and represents an important step for ensuring high quality care for all.

A variety of analytic approaches potentially could be useful for identifying disparities. Performance on a measure could be analyzed by key sociodemographic variables at different levels of analysis such as clinician, facility, or population. As noted above, indirect standardization is based on identifying various categories that could be examined by population and healthcare unit. Multivariable statistical model analysis can provide information about the strength of association of specific factors and how much additional variation in an outcome is accounted for by the variable. However, the Expert Panel recommended stratification as defined in this report to identify disparities and opportunities to reduce disparities.

The Expert Panel did not identify how best to operationalize the use of stratified performance data to identify and reduce disparities. Stratum-specific rates for each unit could prove useful to providers, plans, policymakers, researchers, and the public. However, mechanisms for making detailed data available do not widely exist. How to move toward meaningful use of data and shared accountability for identifying and reducing disparities is a topic that a standing Disparities Committee could address.

Healthcare units need to know whether their performance differs between groups based on sociodemographic factors within their own population. Units also might want to know how their performance with certain groups compares to that of other units. Such data also could prove critical in designing and implementing policies, strategies, and/or programs to improve healthcare equity. Policymakers could use such stratified data to inform funding allocation decisions (e.g., payment rates based on the sociodemographic characteristics of the population). Stratified data could also inform funding for targeted programs such as patient navigators, community health workers, improved access to language services, and other programs designed to mitigate disparities.

The Expert Panel discussed the benefits of transparency with stratified results, but did not resolve how best to present the additional detail in addition to sociodemographic-adjusted scores. Some individuals might find stratified data useful to identify which healthcare units would be best for patients similar to themselves (e.g., income, language, race, ethnicity). CMS or other producers of performance reporting

should make such stratified performance data available when feasible and relevant (e.g., through hyperlinks). At a minimum, it should be publicly available through a clear-cut process for interested parties to request such data. Alternatively, the underlying data needed to construct the stratified performance scores for healthcare units could be made available upon request. Some key issues to be resolved include:

- potential confusion if data are reported more than one way
- cautions about reliability when cell sizes become quite small
- how to construct strata and make drill-down data useful given the potential for use of multiple SDS factors

This is clearly an area where more work needs to done and would benefit from a Standing Disparities Committee. Given the direct relevance of stratified performance data to improved healthcare equity, this is an area where payers such as CMS, states, and health plans could take the lead (as some have done).

Payment and Responsibility for Mitigating Effects of Sociodemographic Factors

During its deliberations the Expert Panel identified two related policy concerns — adequate payment to reflect higher intensity of services to disadvantaged populations and responsibility for mitigating the effects of sociodemographic factors. These concerns, briefly described below, extended beyond the scope of this project but have substantial policy implications.

Disadvantaged populations may have needs that require greater resources. Current payment systems better align resources with clinical/medical needs of patients than services to mitigate the effects of sociodemographic factors. This failure to align payment with supportive patient services for disadvantaged patients creates a mismatch between healthcare unit capacity and the needs of the patient population, thereby creating a potential for worse performance. There are some examples of attempts to adjust payments for services provided to address higher need for resources related to sociodemographic factors. Some examples of this type of payment adjustment are 1) hospital payment adjustment disproportionate share (DSH) of certain low income patients (see overview of Medicare hospital payment); and 2) inclusion of Medicaid status in case-mix adjustment for Medicare Advantage plans (see overview of Medicare Advantage payment). It was beyond the scope of this project to address the adequacy of payment adjustments related to sociodemographic factors. Nonetheless, improved alignment between payments for services and the needs of the patient population served by that unit could potentially partly mitigate the negative effect on patient outcomes. Much of the debate about adjusting for sociodemographic factors relates to setting appropriate expectations for investment in care for disadvantaged patients and concerns about which entity should be incentivized to do so.

Some question whether greater payment to address the needs of sociodemographically complex patients would remove the need for SDS adjustment for performance measures. There is some parallel here to clinical factors, where current case-mix payments to healthcare units aim to account for patient morbidity and severity (and thus need for more costly care); however, performance measures are still risk adjusted for clinical complexity. Similarly, if resources targeted to address the needs of sociodemographically complex patients eventually reduce disparities, the effects of SDS on patient outcome and performance measurement will be reduced but likely it would still be necessary to risk adjust for SDS complexity.

Finally, an important related issue is identifying who is responsible for mitigating the effects of sociodemographic factors on health and healthcare and paying for those efforts. Where does healthcare responsibility end and community responsibility begin? Should the costs of language translation be covered by the community (e.g., multipayer consortium or borne by each healthcare unit, perhaps through enhanced payments)? There are notable examples of extraordinary efforts by healthcare units to address sociodemographic factors such as funding hospice beds for terminally ill homeless patients or providing translators for a large number of languages. These types of efforts require resources above and beyond typical healthcare reimbursement. Just as important a question as who is responsible is the question what is the most effective and efficient approach to address social determinants of health.

Full discussion and resolution of the related issues of payment and responsibility for mitigating the effects were beyond the scope of the Expert Panel's charge, but the recommendations represent a widely-held view among Panel members that improving equity in outcomes will require greater investments.

References

1. Agency for Healthcare Research and Quality (AHRQ). *2012 National Health Care Disparities Report.* Washington, DC: AHRQ; 2013. Report No.: 13-0003. Available at http://www.ahrq.gov/research/findings/nhqrdr/index.html.
2. IOM Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care,* Washington, DC: The National Academies Press; 2003.
3. Schoen C, Radley D, Riley P, et al. *Health Care in the Two Americas: Findings From the Scorecard on State Health System Performance for Low-Income Populations, 2013.* New York, NY: The Commonwealth Fund; 2013. Report No.: Pub No. 1700. Available at http://www.commonwealthfund.org/Publications/Fund-Reports/2013/Sep/Low-Income-Scorecard.aspx?page=all.
4. Satin DJ. Paying physicians and protecting the poor. *Minn Med,* 2006;89(4):42-44.
5. Iezzoni LI. Risk Adjustment for Measuring Health Care Outcomes. 2013;4th[Chicago, IL]: Health Administration Press;
6. Muennig P, Fiscella K, Tancredi D, et al. The relative health burden of selected social and behavioral risk factors in the United States: implications for policy. *American journal of public health,* 2010;2009/12/19:1758-1764.
7. Meyer PA, Yoon PW, Kaufmann RB. Introduction: CDC Health Disparities and Inequalities Report - United States, 2013. *Morbidity and mortality weekly report Surveillance summaries (Washington, D C : 2002),* 2013;2013/11/23:3-5.
8. Adler NE, Newman K. Socioeconomic disparities in health: pathways and policies. *Health Aff (Millwood),* 2002;21(2):60-76.
9. Franks P, Fiscella K. Effect of patient socioeconomic status on physician profiles for prevention, disease management, and diagnostic testing costs. *Med Care,* 2002;40(8):717-724.
10. AHRQ. *2012 National Healthcare Disparities Report* *Http://Www.Ahrq.Gov/Research/Findings/Nhqrdr/Index.Html,* 2014.
11. Asch SM, Kerr EA, Keesey J, et al. Who is at greatest risk for receiving poor-quality health care? *N Engl J Med,* 2006;354(11):1147-1156.
12. van RM, Saha S. Exploring unconscious bias in disparities research and medical education. *JAMA,* 2011;306(9):995-996.
13. Dovidio JF, Fiske ST. Under the radar: how unexamined biases in decision-making processes in clinical interactions can contribute to health care disparities. *Am J Public Health,* 2012;102(5):945-952.
14. Bach PB, Pham HH, Schrag D, et al. Primary care physicians who treat blacks and whites. *N Engl J Med,* 2004;351(6):575-584.
15. Reschovsky JD, O'Malley AS. Do primary care physicians treating minority patients report problems delivering high-quality care? *Health Aff (Millwood),* 2008;27(3):w222-w231.
16. Skinner J, Chandra A, Staiger D, et al. Mortality after acute myocardial infarction in hospitals that disproportionately treat black patients. *Circulation,* 2005;112(17):2634-2641.
17. Chien AT, Wroblewski K, Damberg C, et al. Do physician organizations located in lower socioeconomic status areas score lower on pay-for-performance measures? *J Gen Intern Med,* 2012;27(5):548-554.
18. Fiscella K, Franks P. Influence of patient education on profiles of physician practices. *Ann Intern Med,* 1999;131(10):745-751.
19. Hasnain-Wynia R, Baker DW, Nerenz D, et al. Disparities in health care are driven by where minority patients seek care: examination of the hospital quality alliance measures. *Arch Intern Med,* 2007;167(12):1233-1239.
20. Hong CS, Atlas SJ, Chang Y, et al. Relationship between patient panel characteristics and primary care physician clinical performance rankings. *JAMA,* 2010;304(10):1107-1113.
21. Zaslavsky AM, Hochheimer JN, Schneider EC, et al. Impact of sociodemographic case mix on the HEDIS measures of health plan quality. *Med Care,* 2000;38(10):981-992.

22. Nuzum R, McCarthy D, Gauthier A, et al. *Denver Health: A High-Performance Public Health Care System*. New York, NY: The Commonwealth Fund; 2007. Available at http://www.commonwealthfund.org/Publications/Fund-Reports/2007/Jul/Denver-Health--A-High-Performance-Public-Health-Care-System.aspx.
23. Casalino LP, Elster A. Will Pay-For-Performance And Quality Reporting Affect Health Care Disparities? 2007;26:405-414.
24. Friedberg MW, Safran DG, Coltin K, et al. Paying For Performance In Primary Care: Potential Impact On Practices And Disparities. 2010;29:926-932.
25. Ryan AM. Will Value-Based Purchasing Increase Disparities in Care? 2013;369:2472-2474.
26. Weinick RM, Chien AT, Rosenthal MB, et al. Hospital executivesâ€™ perspectives on pay-for-performance and racial/ethnic disparities in care. 2010;67:574-589.
27. Werner RM, Goldman LE, Dudley RA. Comparison of change in quality of care between safety-net and non-safety-net hospitals. 2008;299:2180-2187.
28. Joynt KE, Jha AK. Characteristics of hospitals receiving penalties under the Hospital Readmissions Reduction Program. *JAMA,* 2013;309(4):342-343.
29. Young GJ, Rickles NM, Chou CH, et al. Socioeconomic characteristics of enrollees appear to influence performance scores for medicare part d contractors. *Health Aff (Millwood),* 2014;33(1):140-146.
30. Zaslavsky AM, Epstein AM. How patients' sociodemographic characteristics affect comparisons of competing health plans in California on HEDIS quality measures. *Int J Qual Health Care,* 2005;17(1):67-74.
31. Werner RM, Asch DA, Polsky D. Racial profiling: the unintended consequences of coronary artery bypass graft report cards. *Circulation,* 2005;111(10):1257-1263.
32. Haider AH, Pronovost PJ. Health information technology and the collection of race, ethnicity, and language data to reduce disparities in quality of care. *Joint Commission journal on quality and patient safety / Joint Commission Resources,* 2011;2011/10/22:435-436.
33. Kuhlthau K, Ferris TG, Iezzoni LI. Risk adjustment for pediatric quality indicators. *Pediatrics,* 2004;113(1 Pt 2):210-216.
34. Zaslavsky AM. Statistical issues in reporting quality data: small samples and casemix variation. *Int J Qual Health Care,* 2001;13(6):481-488.
35. Rothman KG, Greenland S, Lash TL. *Modern Epidemiology,* Lippincott Williams & Wilkins; 2008.
36. Medicare Payment Advisory Commission (MedPAC). *Refining the Hospital Readmissions Reduction Program.* Washington, DC: MedPAC; 2013. Available at http://www.medpac.gov/.
37. Weissman JS, Stern RS, Epstein AM. The impact of patient socioeconomic status and other social factors on readmission: a prospective study in four Massachusetts hospitals. *Inquiry,* 1994;31(2):163-172.
38. Krieger N, Chen JT, Waterman PD, et al. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter?: the Public Health Disparities Geocoding Project. 2002;156:471-482.
39. Link BG, Phelan J. Social conditions as fundamental causes of disease. 1995;Spec No:80-94.
40. Phelan JC, Link BG, Tehranifar P. Social Conditions as Fundamental Causes of Health Inequalities Theory, Evidence, and Policy Implications. 2010;51:S28-S40.
41. Shavers VL. Measurement of socioeconomic status in health disparities research. 2007;99:1013.
42. Dubay LC, Lebrun LA. Health, behavior, and health care disparities: disentangling the effects of income and race in the United States. *International journal of health services : planning, administration, evaluation,* 2012;2013/02/02:607-625.
43. McGrail KM, van Doorslaer E, Ross NA, et al. Income-related health inequalities in Canada and the United States: a decomposition analysis. *American journal of public health,* 2009;2009/01/20:1856-1863.
44. Ngo-Metzger Q, Sorkin DH, Billimek J, et al. The effects of financial pressures on adherence and glucose control among racial/ethnically diverse patients with diabetes. *Journal of general internal medicine,* 2012;2011/10/19:432-437.
45. Kirst M, Shankardass K, Bomze S, et al. Sociodemographic data collection for health equity measurement: a mixed methods study examining public opinions. *International journal for equity in health,* 2013;2013/10/15:75.

46. Davern M, Rodin H, Beebe TJ, et al. The effect of income question design in health surveys on family income, poverty and eligibility estimates. *Health services research,* 2005;2005/09/22:1534-1552.
47. Bonito AJ. *Creation of New Race-Ethnicity Codes and Socioeconomic Status (Ses) Indicators for Medicare Beneficiaries. Final Report, Sub-Task 2.* 2008.
48. Lovasi GS, Weiss JC, Hoskins R, et al. Comparing a single-stage geocoding method to a multi-stage geocoding method: how much and where do they disagree? *International journal of health geographics,* 2007;2007/03/21:12.
49. Goldberg DW, Cockburn MG. Improving Geocode Accuracy with Candidate Selection Criteria. 2010;14:149-176.
50. Sonderman JS, Mumma MT, Cohen SS, et al. A multi-stage approach to maximizing geocoding success in a large population-based cohort study through automated and interactive processes. 2012;6:273.
51. Conti G, Heckman J, Urzua S. The education-health gradient. 2010;100:234-238.
52. Cutler DM, Lleras-Muney A. Understanding differences in health behaviors by education. *Journal of health economics,* 2010;2009/12/08:1-28.
53. Kim M, Zaslavsky AM, Cleary PD. Adjusting Pediatric Consumer Assessment of Health Plans Study (CAHPS) Scores to Ensure Fair Comparison of Health Plan Performances. *Med Care,* 2005;43(1):44-52.
54. O'Malley AJ, Zaslavsky AM, Elliott MN, et al. Case-mix adjustment of the CAHPS Hospital Survey. *Health Serv Res,* 2005;40(6 Pt 2):2162-2181.
55. National Committee on Vital and Health Statistics. Core Health Data Elements, Report of the National Committee on Vital and Health Statistics. National Committee on Vital and Health Statistics, United States Department of health and Human Services, 1996;March 2014.
56. Jorgensen S, Thorlby R, Weinick RM, et al. Responses of Massachusetts hospitals to a state mandate to collect race, ethnicity and language data from patients: a qualitative study. *BMC health services research,* 2010;2011/01/05:352.
57. Escarce JJ, Carreon R, Veselovskiy G, et al. Collection of race and ethnicity data by health plans has grown substantially, but opportunities remain to expand efforts. *Health affairs (Project Hope),* 2011;2011/10/07:1984-1991.
58. Roblin DW. Validation of a neighborhood SES index in a managed care organization. *Medical care,* 2013;2012/03/23:e1-e8.
59. Kwok RK, Yankaskas BC. The use of census data for determining race and education as SES indicators: a validation study. 2001;11:171-177.
60. Roelfs DJ, Shor E, Davidson KW, et al. Losing life and livelihood: a systematic review and meta-analysis of unemployment and all-cause mortality. 2011;72:840-854.
61. Dupre ME, George LK, Liu G, et al. The cumulative effect of unemployment on risks for acute myocardial infarction. 2012;172:1731-1737.
62. Rosenthal L, Carroll-Scott A, Earnshaw VA, et al. The importance of full-time work for urban adults' mental and physical health. 2012;75:1692-1696.
63. Bolton KL, Rodriguez E. Smoking, drinking and body weight after re-employment: does unemployment experience and compensation make a difference? 2009;9:77.
64. Shavers VL. Measurement of socioeconomic status in health disparities research. *J Natl Med Assoc,* 2007;99(9):1013-1023.
65. Galobardes B, Shaw M, Lawlor DA, et al. Indicators of socioeconomic position (part 2). 2006;60:95-101.
66. Wisnivesky JP, Krauskopf K, Wolf MS, et al. The association between language proficiency and outcomes of elderly patients with asthma. 2012;109:179-184.
67. Wisnivesky JP, Kattan M, Evans D, et al. Assessing the relationship between language proficiency and asthma morbidity among inner-city asthmatics. 2009;47:243-249.
68. Schenker Y, Wang F, Selig SJ, et al. The impact of language barriers on documentation of informed consent at a hospital with on-site interpreter services. 2007;22 Suppl 2:294-299.
69. Schenker Y, Karter AJ, Schillinger D, et al. The impact of limited English proficiency and physician language concordance on reports of clinical interactions among patients with diabetes: the DISTANCE study. 2010;81:222-228.

70. McClellan SR, Wu FM, Snowden LR. The impact of threshold language assistance programming on the accessibility of mental health services for persons with limited English proficiency in the Medi-Cal setting. 2012;50:554-558.
71. de Bocanegra HT, Rostovtseva D, Cetinkaya M, et al. Quality of reproductive health services to limited English proficient (LEP) patients. 2011;22:1167-1178.
72. Scheier DB. Barriers to health care for people with hearing loss: a review of the literature. 2009;40:4-10.
73. Hadler MW, Chen X, Gonzalez E, et al. Limited English proficient HMO enrollees remain vulnerable to communication barriers despite language assistance regulations. 2013;1-8.
74. Schyve PM. Language differences as a barrier to quality and safety in health care: the Joint Commission perspective. *Journal of general internal medicine,* 2007;2007/11/28:360-361.
75. Ulmer C, McFadden B, Nerenz DR. *Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement,* National Academies Press; 2009.
76. Nerenz DR, Carreon R, Veselovskiy G. Race, ethnicity, and language data collection by health plans: findings from 2010 AHIPF-RWJF survey. *Journal of health care for the poor and underserved,* 2013;2013/11/05:1769-1783.
77. Bhalla R, Yongue BG, Currie BP. Standardizing race, ethnicity, and preferred language data collection in hospital information systems: results and implications for healthcare delivery and policy. 2012;34:44-52.
78. Lawson EH, Carreon R, Veselovskiy G, et al. Collection of language data and services provided by health plans. *The American journal of managed care,* 2011;2012/01/06:e479-e487.
79. DeNavas-Walt C, Proctor B, Smith J. Current population reports, P60â€“243, Income, poverty and health insurance coverage in the United States: 2011â€“Current population reports. 2012.
80. McWilliams JM, Zaslavsky AM, Meara E, et al. Impact of Medicare coverage on basic clinical services for previously uninsured adults. 2003;290:757-764.
81. McWilliams JM, Zaslavsky AM, Meara E, et al. Health Insurance Coverage And Mortality Among The Near-Elderly. 2004;23:223-233.
82. Abdullah F, Zhang Y, Lardaro T, et al. Analysis of 23 million US hospitalizations: uninsured children have higher all-cause in-hospital mortality. 2010;32:236-244.
83. Wilper AP, Woolhandler S, Lasser KE, et al. Hypertension, Diabetes, And Elevated Cholesterol Among Insured And Uninsured U.S. Adults. 2009;28:w1151-w1159.
84. Kogan MD, Newacheck PW, Honberg L, et al. Association between underinsurance and access to care among children with special health care needs in the United States. *Pediatrics,* 2005;116(5):1162-1169.
85. Lavarreda SA, Brown ER, Bolduc CD. Underinsurance in the United States: an interaction of costs to consumers, benefit design, and access to care. *Annu Rev Public Health,* 2011;32:471-482.
86. Voorhees K, Fernald DH, Emsermann C, et al. Underinsurance in primary care: a report from the State Networks of Colorado Ambulatory Practices and Partners (SNOCAP). *J Am Board Fam Med,* 2008;21(4):309-316.
87. Jean-Jacques M, Persell SD, Hasnain-Wynia R, et al. The implications of using adjusted versus unadjusted methods to measure health care disparities at the practice level. *American journal of medical quality : the official journal of the American College of Medical Quality,* 2011;2011/05/26:491-501.
88. Lurie N, Fremont A. Looking forward: cross-cutting issues in the collection and use of racial/ethnic data. *Health services research,* 2006;2006/08/11:1519-1533.
89. Shavers VL, Fagan P, Jones D, et al. The State of Research on Racial/Ethnic Discrimination in The Receipt of Health Care. 2012;102:953-966.
90. Culhane JF, Goldenberg RL. Racial disparities in preterm birth. *Seminars in perinatology,* 2011;2011/07/30:234-239.
91. Herman WH, Ma Y, Uwaifo G, et al. Differences in A1C by Race and Ethnicity Among Patients With Impaired Glucose Tolerance in the Diabetes Prevention Program. 2007;30:2453-2457.
92. Davidson MB, Schriger DL. Effect of age and race/ethnicity on HbA1c levels in people without known diabetes mellitus: Implications for the diagnosis of diabetes. 2010;87:415-421.

93. Hunt KJ, Gebregziabher M, Lynch CP, et al. Impact of diabetes control on mortality by race in a national cohort of veterans. 2013;23:74-79.
94. Kressin NR, Orner MB, Manze M, et al. Understanding Contributors to Racial Disparities in Blood Pressure Control. 2010;3:173-180.
95. Bosworth HB, Dudley T, Olsen MK, et al. Racial differences in blood pressure control: potential explanatory factors. 2006;119:70-15.
96. Weissman JS, Hasnain-Wynia R. Advancing Health Care Equity through Improved Data Collection. 2011;364:2276-2277.
97. Wynia MK, Ivey SL, Hasnain-Wynia R. Collection of data on patients' race and ethnic group by physician practices. 2010;362:846-850.
98. Kushel MB, Vittinghoff E, Haas JS. Factors associated with the health care utilization of homeless persons. *JAMA : the journal of the American Medical Association,* 2001;2001/02/15:200-206.
99. Baggett TP, O'Connell JJ, Singer DE, et al. The Unmet Health Care Needs of Homeless Adults: A National Study. 2010;100:1326-1333.
100. Doran KM, Ragins KT, Iacomacci AL, et al. The revolving hospital door: hospital readmissions among patients who are homeless. *Medical care,* 2013;2013/08/10:767-773.
101. Baggett TP, Hwang SW, O'Connell JJ, et al. Mortality among homeless adults in Boston: shifts in causes of death over a 15-year period. *JAMA internal medicine,* 2013;2013/01/16:189-195.
102. Greysen SR, Allen R, Rosenthal MS, et al. Improving the quality of discharge care for the homeless: a patient-centered approach. *Journal of health care for the poor and underserved,* 2013;2013/06/04:444-455.
103. U.S. Department of Housing and Urban Development. *Homeless Assistance.* 2009. Available at http://portal.hud.gov/hudportal/HUD?src=/program_offices/comm_planning/homeless . Last accessed March 2014.
104. Stimpson JP, Wilson FA. Cholesterol screening by marital status and sex in the United States. 2009;6:A55.
105. Lee S, Cho E, Grodstein F, et al. Effects of marital transitions on changes in dietary and other health behaviours in US women. 2005;34:69-78.
106. Eng PM, Kawachi I, Fitzmaurice G, et al. Effects of marital transitions on changes in dietary and other health behaviours in US male health professionals. 2005;59:56-62.
107. Hughes ME, Waite LJ. Marital Biography and Health at Mid-Life. 2009;50:344-358.
108. Shor E, Roelfs DJ, Bugyi P, et al. Meta-analysis of marital dissolution and mortality: reevaluating the intersection of gender and age. *Social science & medicine (1982),* 2012;2012/04/27:46-59.
109. Shor E, Roelfs DJ, Curreli M, et al. Widowhood and mortality: a meta-analysis and meta-regression. *Demography,* 2012;2012/03/20:575-606.
110. Kutner M, Greenberg E, Jin Y, et al. *The Health Literacy of America's Adults: Results From the 2003 National Assessment of Adult Literacy.* Washington, DC: Institute of Education Sciences, Natinal Center of Education Statistics, U.S. Department of Education; 2006. Report No.: NCES 2006483. Available at https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006483. Last accessed March 2014.
111. Institute of Medicine CoHL. *Health Literacy: A Prescription to End Confusion.* Washington, DC: The National Academies Press; 2004. Available at http://www.iom.edu/Reports/2004/health-literacy-a-prescription-to-end-confusion.aspx .
112. Kaestle CF, Campbell A, Finn JD, Johnson ST, MicKulecky LJ. *Adult Literacy and Education in America,* Washington, D.C.: National Center for Education Statistics; 2001.
113. von WC, Steptoe A, Wolf MS, et al. Health literacy and health actions: a review and a framework from health psychology. 2009;36:860-877.
114. Medicine Io. *Health Literacy a Prescription to End Confusion,* Washington, D.C.: National Academy Press; 2004.
115. Research AfHQa. Health Literacy Universal Precautions Toolkit. 2011.
116. Wallston KA, Cawthon C, McNaughton CD, et al. Psychometric properties of the Brief Health Literacy Screen in clinical practice. 2014;29:119-126.

117. Chaix B, Rosvall M, Lynch J, et al. Disentangling contextual effects on cause-specific mortality in a longitudinal 23-year follow-up study: impact of population density or socioeconomic environment? 2006;35:633-643.
118. Ponce NA. Measuring Neighborhood Effects and the Use of Geo-coded Variables. 2013.
119. Butler DC, Petterson S, Phillips RL, et al. Measures of social deprivation that predict health care access and need within a rational area of primary care service delivery. *Health services research,* 2013;2012/07/24:539-559.
120. Fremont AM, Bierman AS, Wickstrom SL, et al. Use of Indirect Measures of Race/Ethnicity and Socioeconomic Status in Managed Care Settings to Identify Disparities in Cardiovascular and Diabetes Care Quality. 2005;24:526.
121. Lurie N, Fremont A, Somers SA, et al. The National Health Plan Collaborative to Reduce Disparities and Improve Quality. 2008;34:256-265.
122. Pardo-Crespo MR, Narla NP, Williams AR, et al. Comparison of individual-level versus area-level socioeconomic measures in assessing health outcomes of children in Olmsted County, Minnesota. 2013;67:305-310.
123. Chang CM, Su YC, Lai NS, et al. The combined effect of individual and neighborhood socioeconomic status on cancer survival rates. *PloS one,* 2012;2012/09/08:e44325.
124. Winkleby M, Cubbin C, Ahn D. Effect of cross-level interaction between individual and neighborhood socioeconomic status on adult mortality rates. *American journal of public health,* 2006;2006/11/02:2145-2153.
125. Diez-Roux AV, Kiefe CI, Jacobs DR, Jr., et al. Area characteristics and individual-level socioeconomic position indicators in three population-based epidemiologic studies. 2001;11:395-405.
126. Hawkley LC, Cacioppo JT. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. 2010;40:218-227.
127. Mookadam F, Arthur HM. Social support and its relationship to morbidity and mortality after acute myocardial infarction: Systematic overview. 2004;164:1514-1518.
128. Hibbard JH, Greene J, Becker ER, et al. Racial/Ethnic Disparities And Consumer Activation In Health. 2008;27:1442-1453.
129. Pampel FC, Krueger PM, Denney JT. Socioeconomic Disparities in Health Behaviors. *Annual review of sociology,* 2010;2011/09/13:349-370.
130. Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. 1997;38:21-37.
131. National Quality Forum N. *Healthcare Disparities and Cultural Competency Consensus Standards: Disparities-Sensitive Measure Assessment.* Washington, DC: National Quality Forum; 2012. Available at https://www.qualityforum.org/Publications/2012/11/Healthcare_Disparities_and_Cultural_Competency_Consensus_Standards__Disparities-Sensitive_Measure_Assessment.aspx.
132. National Quality Forum. *Guidance for Evaluating Usability and Use of Performance Measures.* Washington, DC: NQF; 2012. Available at http://www.qualityforum.org/docs/measure_evaluation_criteria.aspx.
133. *A Dictionary of Epidemiology*. 5th ed., Osford Univeristy Press; 2008.
134. National Quality Forum N. *Healthcare Disparities and Cultural Competency Consensus Standards: Technical Report.* Washington, DC: NQF; 2012. Available at https://www.qualityforum.org/projects/Healthcare_Disparities_and_Cultural_Competency.aspx.
135. National Quality Forum. *Patient Reported Outcomes (PROs)in Performance Measurement* . Washington, DC: NQF; 2013. Available at http://www.qualityforum.org/Projects/n-r/Patient-Reported_Outcomes/Patient-Reported_Outcomes.aspx.
136. Braveman PA, Kumanyika S, Fielding J, et al. Health disparities and health equity: the issue is justice. *Am J Public Health,* 2011;101 Suppl 1:S149-55. doi: 10.2105/AJPH.2010.300062. Epub;%2011 May 6.:S149-S155.
137. Austin PC, Alter DA, Tu JV. The use of fixed- and random-effects models for classifying hospitals as mortality outliers: a Monte Carlo assessment. *Med Decis Making,* 2003;23(6):526-539.

Appendix A: Expert Panel on Risk Adjustment for Sociodemographic Factors and NQF Staff

Expert Panel Roster

Kevin Fiscella, MD, MPH (Co-Chair)
Professor, Family Medicine, Public Health Sciences, Community Health and Oncology
University of Rochester Medical Center
Rochester, NY

David Nerenz, PhD (Co-Chair)
Director, Center for Health Policy & Health Services Research
Director of Outcomes Research for the Neuroscience Institute
Vice-Chair for Research of the Department of Neurosurgery at Henry Ford Hospital
Henry Ford Health System
Detroit, MI

Jean Accius, PhD, PMP
Director, Health and Long-Term Services and Supports
AARP
Washington, DC

Alyce Adams, MPP, PhD
Research Scientist II
Chief of Health Care Delivery and Policy, Division of Research
Kaiser Permanente
Oakland, CA

Mary Barger, PhD, MPH, CNM, FACNM
Associate Professor of Nursing, University of California San Diego
San Diego, CA

Susannah M. Bernheim, MD, MHS
Director of Quality Measurement
Yale New Haven Health System Center for Outcomes Research and Evaluation (CORE)
New Haven, CT

Monica Bharel, MD, MPH
Chief Medical Officer
Boston Health Care for the Homeless Program
Boston, MA

Mary Beth Callahan, ACSW/LCSW
Senior Social Worker
Dallas Transplant Institute
Dallas, TX

Lawrence Casalino, MD, PhD
Livingston Farrand Professor of Public Health
Chief, Division of Outcomes and Effectiveness Research
New York, NY

Alyna Chien, MD, MS
Assistant Professor, Pediatrics, Harvard Medical School
Pediatrician, Children's Hospital Primary Care Center, Boston Children's Hospital
Boston, MA

Marshall Chin, MD, MPH
Richard Parrillo Family Professor of Healthcare Ethics in the Department of Medicine,
University of Chicago
Director of the RWJF Finding Answers: Disparities Research for Change National Program Office
Chicago, IL

Mark Cohen, PhD
Statistical Manager, Continuous Quality Improvement, Division of Research and Optimal Patient Care
American College of Surgeons
Adjunct Associate Professor of Surgery, Feinberg School of Medicine, Northwestern University
Chicago, IL

Nancy Garrett, PhD
Chief Analytics Officer
Hennepin County Medical Center
Minneapolis, MN

Norbert Goldfield, MD
Medical Director, 3M HIS Clinical and Economics Research**,** 3M
Staff Physician, Brightwood Community Health Center
Wallingford, CT

Atul Grover, MD, PhD, FCCP
Chief Public Policy Officer
Association of American Medical Colleges
Washington, DC

David Hopkins, PhD
Senior Advisor**,** Pacific Business Group on Health
Adjunct Associate, Center for Primary Care and Outcomes Research and Center for Health Policy,
Stanford University
San Francisco, CA

Dionne Jimenez, MPP
Research & Policy Coordinator
Service Employees International Union
Los Angeles, CA

Steven Lipstein, MHA
President and CEO
BJC HealthCare
St. Louis, MO

Eugene Nuccio, PhD
Assistant Professor, Division of Health Care Policy and Research, School of Medicine
University of Colorado, Anschutz Medical Campus
Aurora, CO

Sean O'Brien, PhD
Assistant Professor, Biostatistics and Bioinformatics
Duke University Medical Center
Durham, NC

Pam Owens, PhD
Senior Research Scientist and Scientific Director of the AHRQ Quality Indicators™
AHRQ
Rockville, MD

Ninez Ponce, MPP, PhD
Professor, Department of Health Policy and Management
UCLA Fielding School of Public Health
Los Angeles, CA

Thu Quach, PhD, MPH
Research Director
Asian Health Services
Oakland, CA

Tia Goss Sawhney, DrPH, FSA, MAAA
Director of Data, Analytics, and Research
Illinois Department of Healthcare and Family Services
Chicago, IL

Nancy Sugg, MD, MPH
Medical Director
Pioneer Square Clinic & Downtown Homeless Programs, Harborview Medical Center
Associate Professor of Medicine, Division of General Internal Medicine at the University of Washington
Seattle, WA

Rachel Werner, MD, PhD
Associate Professor of Medicine
University of Pennsylvania
Philadelphia, PA

Expert Panel Member Biographies

Kevin Fiscella, MD, MPH (Co-Chair)
Professor, Family Medicine, Public Health Sciences, Community Health and Cancer Center, University of Rochester Medical Center
Kevin Fiscella, MD, MPH, is a tenured Professor of Family Medicine, Public Health Sciences, and Community Health, and Oncology at the University Rochester School Medicine. His scholarly work has focused on both conceptual models and empirical research related to health care disparities. His current work addresses practical strategies to mitigate disparities in health care quality. He has served on numerous national committees related to health and health care disparities. He has published more than 180 papers in peer-reviewed journals and has received major research grants from numerous federal agencies and private foundations.

David Nerenz, PhD (Co-Chair)
Director, Center for Health Policy & Health Services Research, Henry Ford Health System
David R. Nerenz, PhD, is Director of the Center for Health Policy and Health Services Research at Henry Ford Health System in Detroit. He is also Director of Outcomes Research for the Neuroscience Institute, and Vice-Chair for Research of the Department of Neurosurgery at Henry Ford Hospital. He was appointed in May of 2012 as a Commissioner on the Medicare Payment Advisory Commission (MedPAC). He recently served as the Chair of the Institute of Medicine Committee on Leading Health Indicators for Healthy People 2020 and Chair of the IOM Subcommittee on Standardized Collection of Race/Ethnicity Data for Healthcare Quality Improvement.

Jean Accius, PhD, PMP
Director, Health and Long-Term Services and Supports, AARP
Jean Accius, PhD, is an expert in health and long-term care policy. His background includes translating research into policy and practice. Currently, Jean is the Director of Health and Long-term Services and Supports (LTSS) at AARP. In this capacity, he leads the policy development process on health and LTSS related issues that guides AARP's legislative, regulatory and litigation activities. He also provides strategic advice and counsel to senior AARP leadership and other departments to ensure policy integration and consistency across the association.

Alyce Adams, MPP, PhD
Research Scientist II and Chief of Health Care Delivery and Policy, Division of Research, Kaiser Permanente
Alyce S. Adams, MPP, PhD, is Research Scientist II and Chief of Health Care Delivery and Policy at the Kaiser Permanente Division of Research in Oakland, California. Her research explores disparities in chronic disease treatment outcomes using longitudinal data methods. Dr. Adams' current studies include a cluster randomized clinical trial to improve diabetic peripheral neuropathy treatment outcomes funded by the Patient Centered Outcomes Research Institute and an evaluation of the impact of Medicare Part D among dual Medicaid and Medicare enrollees funded by the National Institute on Aging. She has a PhD in Health Policy from Harvard University.

Mary Barger, PhD, MPH, CNM, FACNM

Associate Professor of Nursing, University of California San Diego, and American College of Nurse Midwives

Mary Barger PhD, MPH, CNM, FACNM, is a perinatal epidemiologist and has practiced clinical nurse-midwifery for over 25 years. She has taught in a school of public health, a medical school, and two schools of nursing. The focus of her research has been on maternal morbidity and mortality with a focus on cesareans and using administratively collected data, such as hospital discharge data and birth data, to examine processes and outcomes of care. One of her recent studies combined survey data with GIS information to further understand the racial disparities in cesarean rates in California

Susannah M. Bernheim, MD, MHS

Director, Quality Measurement, Yale New Haven Health System Center for Outcomes Research and Evaluation (CORE)

Susannah M. Bernheim, MD, MHS, is a Family Physician, Health Services Researcher and the Director of Quality Measurement at Yale-New Haven Hospital's Center for Outcomes Research and Evaluation (CORE). She has extensive experience leading teams in measure development, maintenance, NQF endorsement, and implementation. Her research focuses on the intersection of healthcare quality, outcomes and socioeconomic status. She received her MD at the University of California San Francisco, and her Master of Health Sciences at Yale University. Following a research fellowship and prior to joining CORE she served as Deputy Director of Performance Management for the Yale New Haven Health System.

Monica Bharel, MD, MPH

Chief Medical Officer, Boston Health Care for the Homeless Program

Monica Bharel, MD, is the Chief Medical Officer for the largest nonprofit health care organization for homeless individuals in the country. Under her leadership, the organization achieved Level 3 NCQA PCMH recognition and a 3-fold improvement in women's health quality indicators. She was appointed by Governor Patrick to serve on the Behavioral Health Integration Task Force under the Massachusetts payment reform initiative. She has spoken locally and nationally about the needs of vulnerable and homeless individuals. She is currently focused on ensuring that state and national health care reform efforts enhance the care for homeless individuals, without inadvertently widening health care disparities.

Mary Beth Callahan, ACSW/LCSW

Senior Social Worker, Dallas Transplant Institute

Mary Beth Callahan has worked in nephrology social work since 1984. She is currently Senior Social Worker at Dallas Transplant Institute (DTI) and has had the privilege to work with ESRD patients on hemodialysis, peritoneal dialysis and transplant. She has served on numerous advisory boards and professional committees. She served as CNSW Chair from 1996-1998 and is co-developer of the CNSW's Outcomes Training Program. Her focus with transplant recipients is to help them prepare to return to work whenever possible and/or to live life to the fullest. One of her ongoing efforts is to encourage other staff members to keep rehabilitation in the forefront of their minds. She hosts Job Club monthly at DTI. Job Club developed from joint research grants from the Society for Transplant Social Workers and the Council of Nephrology Social Workers and provides patients with information on Social Security Work Incentives and connects patients with vocational rehabilitation resources and hope.

Lawrence Casalino, MD, PhD

Livingston Farrand Professor of Public Health; Chief, Division of Outcomes and Effectiveness Research, Weill Cornell Medical College

Lawrence Casalino, MD, PhD, has written some of the seminal articles on unintended consequences of quality measurement and on SES disparities and quality measurement. He has also served on relevant national committees. He has quite a lot of knowledge about the organization of medical practice and hospital care and about the responses of providers to incentives. This knowledge comes from quantitative and qualitative research as well as from the 20 years that he spent as a family physician in full-time practice, during which time he also served as a hospital medical staff president and vice president of a large independent practice association.

Alyna Chien, MD, MS

Assistant Professor, Boston Children's Hospital

Alyna Chien, MD, is a physician health services researcher at Harvard Medical School and Boston Children's Hospital. She is the leading pediatric expert on the use of performance incentives in healthcare and has extensive experience using established risk adjustment methods. Currently, she is examining whether geocoded socioeconomic information can improve pediatric risk adjustment algorithms so that healthcare payments can better reflect pediatric patient complexity. She has used similar geocoding techniques to examine the degree to which socioeconomic factors have affected the ability of very large physician organizations to respond to performance incentives. Her work is funded by AHRQ, NICHD, and RWJF.

Marshall Chin, MD, MPH

Richard Parrillo Family Professor of Healthcare Ethics in the Department of Medicine, University of Chicago

Marshall H. Chin, MD, MPH, FACP, Richard Parrillo Family Professor of Healthcare Ethics in the Department of Medicine at the University of Chicago, is a general internist with extensive experience improving the care of vulnerable patients with chronic disease. Dr. Chin is Director of the RWJF Finding Answers: Disparities Research for Change National Program Office. He was a member of the IOM Committee on Future Directions for the National Healthcare Quality and Disparities Reports. He serves on the NQF MAP Coordinating Committee and was a member of the NQF Healthcare Disparities and Cultural Competency Consensus Standards Steering Committee.

Mark Cohen, PhD

Statistical Manager, Continuous Quality Improvement, Division of Research and Optimal Patient Care, American College of Surgeons

Mark Cohen, PhD, is the Statistical Manager, Continuous Quality Improvement, Division of Research and Optimal Patient Care, American College of Surgeons, and Adjunct Associate Professor of Surgery, Feinberg School of Medicine, Northwestern University. Since 2008, he has managed statistical efforts related to the ACS National Surgical Quality Improvement Program (NSQIP), ACS Universal Surgical Risk Calculator, ACS Metabolic and Bariatric Surgery Accreditation and Quality Improvement Program (MBSAQIP), and the ACS NSQIP Pediatric program. He has 125 publications and his current research focuses on optimizing risk-adjustment and reporting methodologies used in these programs. Before joining the ACS, Dr. Cohen was Statistician and, later, Technical Director at the Naval Institute for Dental and Biomedical Research.

Nancy Garrett, PhD

Chief Analytics Officer, Hennepin County Medical Center

Nancy Garrett, PhD, is currently Chief Analytics Officer at Hennepin County Medical Center, where she is developing methods to measure the impact of socioeconomic status on cost and quality measures for HCMC's diverse safety net population. She has an extensive background in applied health services research, and authored a chapter on provider profiling in a managed care textbook. Nancy is on NQF's Cost and Resource Use Steering Committee where she raised issues about adjusting for socioeconomic status that helped lead to the convening of this expert panel. Nancy has a Ph.D. in Demography from the University of Illinois.

Norbert Goldfield, MD

Medical Director, 3M HIS Clinical and Economics Research, 3M

Dr. Goldfield works as a medical director of 3MHIS, developing classification tools linking payment to quality. This work is used throughout the United States and overseas, with public and private payers. Dr. Goldfield is a board certified internist practicing at a community health center. He edits the peer reviewed Journal of Ambulatory Care Management and has published extensively. He is on a number of boards including Health Care for All. He is also the founder and executive director of Healing Across the Divides (www.healingdivides.org), an organization seeking to improve the health of Israelis and Palestinians.

Atul Grover, MD, PhD, FCCP

Chief Public Policy Officer, Association of American Medical Colleges

Atul Grover, MD, PhD, is the Chief Public Policy Officer for the Association of American Medical Colleges (AAMC). In this role, he manages the AAMC's health, educational, and scientific policies. Dr. Grover joined the AAMC in its Center for Workforce Studies, where he managed research activity and directed externally funded workforce studies. Prior to the AAMC, Dr. Grover was a senior consultant in health care finance and applied economics for The Lewin Group, Inc and also served with the Health Resources and Service Administration. Dr. Grover is a clinical faculty member at the George Washington School of Medicine.

David Hopkins, PhD

Senior Advisor, Pacific Business Group on Health

David S. P. Hopkins, PhD, is Senior Advisor at the Pacific Business Group on Health. Hopkins is also affiliated with the Center for Health Policy and the Clinical Excellence Research Center at Stanford University Medical School. He earned his A.B. in Biology from Harvard, and his M.S. in Statistics and Ph.D. in Operations Research from Stanford. Hopkins chaired the California Cooperative Healthcare Reporting Initiative (CCHRI) Executive Committee from 1996-2012 and is the former Chair of the Integrated Healthcare Association Pay-for-Performance Technical Efficiency Committee. He served two terms on NQF's Consensus Standards Approval Committee and currently chairs the NQF Purchaser Council.

Dionne Jimenez, MPP

Research & Policy Coordinator, Service Employees International Union

Dionne Jimenez is a research and policy coordinator for the Service Employees International Union, which represents 2.2 million workers advocating to improve their lives and the services they provide. She performs public policy analysis and develops public policy positions for SEIU on health care

financing, workforce, quality of care & life, and other key issues related to the health care sector. Previous professional experience includes serving as staff and legislative assistant to Congressman George Miller (D-CA). Dionne is a proud first-generation college graduate. She has a Master of Public Policy degree from the University of California, Los Angeles, School of Public Affairs, and a BA in Political Science from the University of California, Berkeley.

Steven Lipstein, MHA
President and CEO, BJC Healthcare
Steven Lipstein has led BJC Healthcare since 1999. He is highly engaged in ensuring that people everywhere receive high-quality, safe care. BJC is the largest provider of uncompensated care in Missouri. From 2008 to 2010, Mr. Lipstein co-chaired the oversight committee for Missouri Medicaid. Prior to joining BJC, Mr. Lipstein held executive roles at the University of Chicago and The Johns Hopkins Health System. He is vice chair of the Board of Governors for the Patient-Centered Outcomes Research Institute. He graduated from Emory University, has an MHA from Duke University, and completed an administrative fellowship at Massachusetts General Hospital.

Eugene Nuccio, PhD
Assistant Professor, University of Colorado, Anschutz Medical Campus
Eugene Nuccio, PhD, Assistant Professor, holds a doctorate in Education Psychology and has extensive experience with statistical analysis, measurement, and risk adjustment of outcomes for Medicare home care recipients. Since 2004 he has led the development of the last three sets of 40+ prediction models used to risk adjust home health outcomes nationally. He initiated innovations in how to represent OASIS data as well as methodological changes to develop complex multivariate models. Under the direction of MedPAC, Dr. Nuccio linked CMS claims, OASIS, and other data sources to produce experimental quality measures and prediction models. His contributions to the scientific literature on risk adjustment include presentations at AcademyHealth.

Sean O'Brien, PhD
Assistant Professor, Biostatistics and Bioinformatics, Duke University Medical Center
Sean O'Brien, PhD, is an Assistant Professor in the Department of Biostatistics and Bioinformatics at Duke University Medical Center. Since 2005, he has served as statistical director of the Society of Thoracic Surgeons (STS) Data Warehouse and Analysis Center and as co-investigator of several grants and contracts using large registries to study comparative effectiveness and health care quality. Dr. O'Brien also works on the development and evaluation of quantitative methods for healthcare provider performance assessment. His research interests include risk adjustment methodology, composite measures, and Bayesian modeling.

Pam Owens, PhD
Senior Research Scientist, AHRQ
Pamela Owens, PhD, is a senior research scientist at the Agency for Healthcare Research and Quality (AHRQ). Dr. Owens is the Scientific Director of the AHRQ Quality IndicatorsTM (QIs) and co-leads Healthcare Cost and Utilization Project (HCUP) outpatient data development. Dr. Owens' research experience includes the quality and access to care for various populations, conditions and settings, including children, low income, mental health, asthma, readmissions, ambulatory surgery, emergency department and inpatient settings. Her work has appeared in journals such as the JAMA, Medical Care, Health Services Research, Annals of Internal Medicine, Pediatrics, Academic Emergency Medicine,

Psychiatric Services, and Journal of Preventive Medicine. Dr. Owens received a Ph.D. in epidemiology and health policy from Yale University and completed a post-doctoral fellowship at Johns Hopkins. She also has six years of clinical experience as an occupational therapist.

Ninez Ponce, MPP, PhD
Professor, Department of Health Policy and Management - UCLA Fielding School of Public Health
Ninez Ponce, MPP, PhD, is a professor in the Department of Health Policy and Management at the UCLA Fielding School of Public Health. In 14 years at UCLA, she has taught courses on health insurance, health economics, health policy, and research methods, with a research focus on racial/ethnic disparities in cancer prevention and control. She also conducted program evaluation, research, and public policy for a W.K. Kellogg Foundation national initiative to improve health care for the underserved. She has served on expert advisory groups for the Institute of Medicine, the Office of the Patient Advocate, and the UCLA Department of Health Services.

Thu Quach, PhD, MPH
Research Director, Asian Health Services
Thu Quach, PhD, MPH, is an epidemiologist and primary research interest has focused on the influence of environmental and sociocultural factors on immigrant population health. As a research scientist at the Cancer Prevention Institute of California, a non-profit research organization, she leads research studies focusing on the booming nail salon workforce, comprised mainly of Vietnamese immigrants. In 2011, after years of research collaboration, she was recruited by Asian Health Services to become the inaugural research director at this community health center (CHC) serving low-income Asian Americans patients. She spearheads efforts across several CHCs to incorporate social determinants of health factors in risk adjustment.

Tia Goss Sawhney, DrPH, FSA, MAAA
Director of Data, Analytics, and Research, Illinois Department of Healthcare and Family Services
Tia Goss Sawhney, DrPH, FSA, MAAA, is the Director of Data, Research, and Analytics with the Illinois Medicaid plan. She is the author of the 2010 paper "Health Insurance Risk Adjustment: The Income Effect". The paper is included in her 2012 dissertation "Controlling Indirect Selection under Healthcare Reform" available at *www.soa.org/files/sections/health-dissertation-**sawhney**.pdf.* She is Fellow of the Society of Actuaries and a Member of the American Academy of Actuaries and active in each organization.

Nancy Sugg, MD, MPH
Medical Director Pioneer Square Clinic & Downtown Homeless Programs, Harborview Medical Center
Nancy Sugg, MD, MPH, is Associate Professor of Medicine in the Division of General Internal Medicine at the University of Washington and Medical Director of Harborview Medical Center's Pioneer Square Clinic and Downtown Homeless Programs. She is the Chair of the Care Management Committee at Harborview Medical Center, focusing on decreasing inpatient lengths of stay and readmissions and improving transitions of care. She works closely with Seattle-King County Public Health's Healthcare for the Homeless Network developing integrated medical services for homeless and outcomes measures for clinics caring for underserved populations. Dr. Sugg directs research projects and mentors future primary care providers and policy makers for underserved populations.

Rachel Werner, MD, PhD

Associate Professor of Medicine, University of Pennsylvania

Rachel Werner, MD, PhD, is an Associate Professor of Medicine at the University of Pennsylvania. She received her medical degree from the University of Pennsylvania School of Medicine, where she also did her residency in Internal Medicine. While completing a clinical fellowship in general internal medicine, she also received a Ph.D. in health economics from the Wharton School at the University of Pennsylvania. Dr. Werner's research seeks to understand the effect of healthcare policies and delivery systems on quality of care. In particular, she has examined the role of provider payment and financial incentives on provider behavior, the organization of healthcare, racial disparities, and overall healthcare quality. Her work has empirically investigated numerous unintended consequences to quality improvement incentives and was among the first to recognize that public reporting of quality information may worsen racial disparities. She is currently principal investigator of an R01 from the Agency of Healthcare Research and Quality (examining how pay-for-performance in hospitals changed the value of healthcare) and an R01 from the National Institute of Aging (examining the effect of Medicaid pay-for-performance for nursing homes on delivery of nursing home care). She also directs one of five national centers to evaluate the effectiveness of the medical home by the Veterans Health Administration. She has received numerous awards including the Dissertation Award and the Alice Hersh New Investigator Award from Academy Health and the Presidential Early Career Award for Scientists and Engineers. Her research has been published in high-impact peer-reviewed journals, including *JAMA*, *Journal of Health Economics, Health Services Research,* and *Health Affairs*. In addition to her research, Dr. Werner is a practicing primary care internist at the Philadelphia VA Medical Center and regularly attends the hospital's internal medicine service. She supervises healthcare provided by Hospital of the University of Penn.

NQF Staff

Helen Burstin, MD, MPH
Chief Scientific Officer

Karen Beckman Pace, PhD, MSN
Senior Director
Performance Measurement

Taroon Amin, MA, MPH
Senior Director
Performance Measurement

Karen Johnson, MS
Senior Director
Performance Measurement

Suzanne Theberge, MPH
Senior Project Manager
Performance Measurement

Erin O'Rourke
Senior Project Manager
Performance Measurement

Appendix B: Glossary

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).[132]

Confounding – The distortion in the degree of association between an exposure (independent variable) and an outcome (dependent variable) due to a mixing of effects between the exposure and an incidental (confounding) factor. Confounding represents systematic error and threatens the internal validity of an epidemiologic study since it can lead to false conclusions regarding the true relationship between an exposure and outcome.

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. [134]

Outcome – the result of providing healthcare. The term outcome will be used to broadly include the following types of outcomes relevant to performance measurement:

- Quality outcomes include:
  - Health outcome is the health status of a patient (or change in health status) resulting from healthcare—desirable or adverse.
  - In some situations, resource use may be considered a proxy for a health state (e.g., hospitalization may represent deterioration in health status).
  - Intermediate clinical outcome is a change in physiologic state that leads to a longer-term health outcome (e.g., hemoglobin, blood pressure).
  - Patient-reported outcome is any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else. The domains of PROs include health-related quality of life/functional status, symptom/symptom burden, experience with care (including engagement, activation), and health-related behaviors.[135]
- Economic outcomes include the cost and resource use associated with providing healthcare services. (Although efficiency is considered one aspect of quality, cost and resource use alone without consideration of quality is not considered a quality performance measure.)

Peer groups for comparison – Creation of peer groups of providers caring for a similar mix of patients, within which to examine performance scores.

Performance measure – Numeric quantification of healthcare quality for a designated accountable entity such as hospital, health plan, nursing home, clinician, etc. (NQF measure testing report)

Risk Adjustment (also known as case-mix adjustment) – Statistical methods to control or account for patient-related factors when computing performance measure scores; methods include multivariable modeling, indirect standardization, or direct standardization. These methods can be used to produce a ratio of observed to expected, a risk-adjusted rate, or other estimate of performance.

Social Determinants of Health – Healthy People 2020 defines social determinants of health as conditions in the environments in which people are born, live, learn, work, play, worship, and age that affect a wide range of health, functioning, and quality-of-life outcomes and risks. Conditions (e.g., social, economic, and physical) in these various environments and settings (e.g., school, church, workplace, and neighborhood) have been referred to as "place." In addition to the more material attributes of "place," the patterns of social engagement and sense of security and well-being are also affected by where people live. Resources that enhance quality of life can have a significant influence on population health outcomes. Examples of these resources include safe and affordable housing, access to education, public safety, availability of healthy foods, local emergency/health services, and environments free of life-threatening toxins.

Social disadvantage – Braveman et al. define social disadvantage as "Unfavorable social, economic, or political conditions that some groups of people systematically experience based on their relative position in social hierarchies.[136] Social disadvantage indicates restricted ability to participate fully in society and enjoy the benefits of progress. Social disadvantage is reflected, for example, by low levels of wealth, income, education, or occupational rank, or by less representation at high levels of political office.

Sociodemographic – Broad term referring to a variety of socioeconomic (e.g., income, education, occupation) and demographic factors (age, race, ethnicity, primary language).

Socioeconomic Status – Broadly conceptualized as one's relative position within society. Socioeconomic status has traditionally been defined and measured by education, income, and occupation.[8]

Stratification – Computing performance scores separately for different strata or groupings of patients based on some characteristic(s) – i.e., each healthcare unit has multiple performance scores (one for each stratum) rather than one overall performance score.

Appendix C: Outcome Performance Measures and Risk Adjustment – the Basics

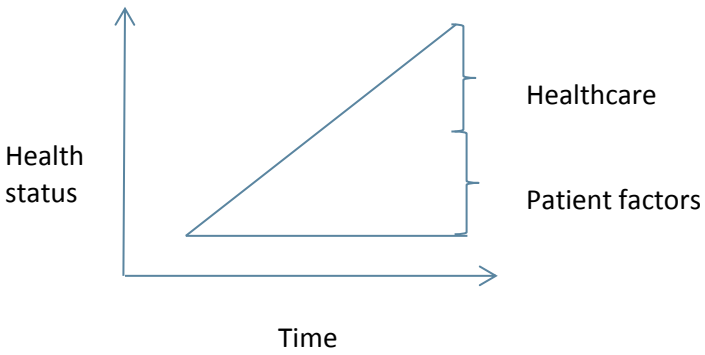
Outcome performance measures aggregate the data on individual patient outcomes for an accountable entity (e.g., hospital, clinician, nursing home). Outcomes generally are a function of several inputs including patient factors, treatment effectiveness, quality of care, and random events. This can be represented as an equation:

> **Outcomes = *f*** (intrinsic patient factors, treatment effectiveness, quality of care, random chance) [5, p. 5]

This equation is a simplified description because outcomes also may be a function of complex interaction among these factors.

Outcomes often represent a change in some health status indicator (e.g., function, pain) over time; that change can be due to both healthcare and patient factors as represented in Figure 1. Some outcomes, such as hospital readmission, are considered a proxy for a change in health status.

Figure C-1. Outcome as a Change over Time



Risk Factors

Iezzoni[5, p. 31] identified the major categories for the potential patient factors that may influence outcomes to include the following. This is not a comprehensive list and concepts may overlap. Additionally, not all factors may affect every outcome.

- Genetics (e.g., predisposition to conditions or health-related behaviors)
- Demographic characteristics (e.g., age, sex, race, ethnicity, primary language)
- Clinical factors (diagnoses, conditions and severity; physiologic stability; physical, mental, cognitive function)
- Psychosocial factors, socioeconomic, and environmental factors (e.g., family, education, occupation, economic resources, health insurance, neighborhood)
- Health-related behaviors and activities (tobacco, diet, physical activity)
- Quality of life, attitudes, and perceptions (health-related quality of life and overall health status; preferences; cultural, religious beliefs and behavior)

The final selection of risk factors involves an iterative process using the guidelines identified in Table 5 including:

- Clinical/conceptual relationship with the outcome of interest
- Empirical association with the outcome of interest
- Variation in prevalence of the factor across the measured entities
- Present at the start of care
- Does not represent the care provided (e.g., treatments, expertise of staff)
- Resistant to manipulation or gaming
- Accurate data that can be reliably and feasibly captured
- Contribution of unique variation in the outcome (not redundant)
- Potentially, improvement in risk model metrics of discrimination and/or calibration
- Potentially, face validity and acceptability

Risk Adjustment in Outcome Performance Measurement

The ultimate goal of performance measurement is to facilitate improvement in healthcare and health. Measurement is used to identify differences in quality of healthcare and identify opportunities for improvement. Unlike many process performance measures, which are focused on care practices that should be delivered to all patients in a specified target population, the goal for outcome performance may not be 100% (or 0%). Due to the limits of science, not all patients will achieve the outcome (e.g., survive), and the "right" rate may not be known. Consequently, it is through comparison across providers that opportunities for improvement are identified. Providers with superior risk-adjusted outcomes set the goal for what is possible to achieve. In order for performance results to be meaningful and valid for identifying differences in performance across providers, outcome performance measures must be adjusted for different levels of risk in the patients served.

Outcome performance measurement is intended to identify the effect of care on the outcome of interest in order to make a conclusion about quality and direct efforts for quality improvement. As indicated in the equation and Figure 1, the relationship between healthcare and the outcome may be confounded by various patient factors. That is, patient factors (e.g., severity or complexity) are also correlated with the outcome and provide an alternative explanation for the outcome. Confounding factors need to be controlled or adjusted for in order to make conclusions about the quality of care based on performance on the outcome measure.

Risk Adjustment refers to statistical methods to control or account for patient-related factors when computing performance measure scores, including multivariable models, indirect standardization, or direct standardization. These methods can be used to produce a ratio of observed to expected and/or a risk-adjusted rate. Risk adjustment refers to the operations performed during the calculation of the performance score. 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 – i.e., what is expected if the hypothetical average provider cared for the specific mix of patients
    - Extension of indirect standardization to multivariable statistical models [5]

- Direct standardization where provider-specific rates are calculated in each stratum and applied to the standard population case mix, producing an estimate of what would be expected if the provider were to treat the standard case mix. [5] This approach is not commonly used to profile performance.

Risk Model Evaluation

Statistical risk models are often evaluated on model discrimination (extent to which the model predicts higher probabilities of the outcome for patients who experienced the outcome than for those who did not) and calibration (the match between predicted and actual outcome rates within subgroups of the data such as risk deciles). It is important to recognize when assessing risk models used for outcome performance measures, the metrics of model discrimination such as C-statistic or R-squared are not necessarily expected to achieve comparable values as models that include and are intended to explain the contribution of all variables that influence the outcome. In risk models, the independent variables are purposely limited to patient risk factors; variables related to care processes or structures are not included so that differences in risk-adjusted outcome rates can be attributed to differences in the care provided, i.e., differences in quality.

Approaches to Statistical Modeling

Statistical modeling to estimate the provider score on the outcome involves choosing from among a variety of options including:

- Random effects with shrinkage estimators vs. fixed effects
- Shrinking toward the overall average or some other benchmark (e.g., average of “like” providers
- Hierarchical models
- Bayesian analysis

The various methods may have different trade-offs and policy implications. For example, fixed effects models identify more outliers, some of which will be false positives; whereas, random effects models identify fewer outliers, some of which will be false negatives.[137]

Appendix D: Confounding – the Basics

Confounding is an epidemiological term that refers to the distortion in the degree of association between an exposure (independent variable) and an outcome (dependent variable) due to a mixing of effects between the exposure and an incidental (confounding) factor.{Porta} Confounding represents systematic error and threatens the internal validity of an epidemiologic study since it can lead to false conclusions regarding the true relationship between an exposure and outcome.

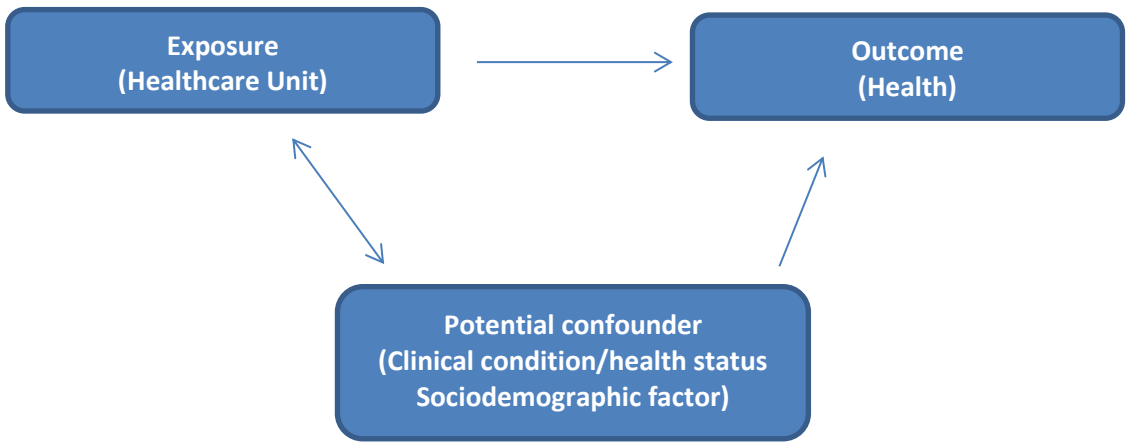
In the field of epidemiology, researchers often are interested in determining whether, how, and to what extent—an "exposure" to a particular entity (e.g., a microbe, a medication, or a procedure) is related to a particular outcome (e.g., a sickness, a recovery, or an improvement). The direction and magnitude of that relationship between the exposure of interest and the outcome is known as the "effect size"; it can be positive or negative, large or small, and statistically significant or not. In the case of outcome performance measurement, the "exposure" of interest is the healthcare unit's structures and processes of care that influence some particular outcome (e.g., mortality).

Usually, however, there are other factors—in addition to the exposure of interest—that are associated with that particular outcome. If such factors are related to the exposure of interest and are causally related to the outcome of interest, they can distort the effect size. This distortion is known as *confounding* and those other factors are known as potential c*onfounders*. The three characteristics of potential confounders are as follows:

- they are a risk factor for the outcome of interest,
- they are associated with the exposure of interest, and
- they are not affected by either the exposure or the outcome.[35]

Importantly, the third characteristic indicates that potential confounders do not represent an intermediate step in the causal pathway between the exposure of interest and the outcome; also, it can be satisfied by factors that precede both the exposure of interest and the outcome. The relationship between the exposure of interest, the outcome of interest, and potential confounders is shown in Figure B-1.

Figure D-1. Relationship between exposure, outcome, and potential confounders



Depending on the strength of the relationships between potential confounders and the exposure and outcome, the type and degree of distortion in the effect size can vary. For example, confounding can make an effect size appear to be statistically significant when it is not (that is, there may appear to be an actual relationship between an exposure of interest and a particular outcome, even when there is not one) or vice-versa. Confounding also can change the direction or magnitude of the effect size (that is, the relationship may appear to be a positive one when in fact it is negative, or it may appear larger [or smaller] than it actually is). Because confounders obscure the relationship between the exposure of interest and the outcome, researchers try to eliminate (or at least minimize) the distortion by "adjusting" for confounding factors in some way (often using statistical techniques).

The discussion of confounders can be extended conceptually to outcome performance measurement. The purpose of outcome performance measurement is to identify the effect of care on health-related outcomes, in order to make a conclusion. In this case, the "exposure" of interest is to the health care unit and its various structures and processes of care that influence some particular outcome (e.g., mortality). As noted above, if other factors are associated with—but not the result of—actual structures or processes used in the provision of care and also influence the outcome of interest, the true "contributions" of the care structures/processes to the outcome may be obscured, because they are "mixed with" or distorted by the contributions of those other factors. In order to make correct conclusions about quality, adjustment for potential confounders is needed. In outcome performance measurement, potential confounders include patient-level characteristics that are risk factors for the outcome of interest that are present prior to the provision of care. Adjustment for such factors is known as risk adjustment or case mix adjustment.

Clinical factors present at the start of care (e.g., severity of illness) and/or other health status factors (e.g., self-reported health) typically are considered potential confounders in outcome performance measurement and therefore are included in risk adjustment strategies. However, risk factors such as genetic characteristics, sociodemographic factors, health-related behaviors, and less commonly available patient-level factors such as beliefs, attitudes, and perceptions may also be potential confounders and if so, should be included in the risk-adjustment strategy, as failure to do so may conceal the true relationship between the structure/processes of care and lead to incorrect conclusions about the quality of that care.

A complication that develops during the course of care can affect the outcome, but should not be considered a confounder because it is in the causal pathway between the exposure to the healthcare unit and the outcome.

Appendix E: Example of Checking for Between-Unit Effect

An example to analyze within- versus between-unit covariate effects when a patient-level variable "low income" is significant is to add to the model a unit-level variable "percent of patients of low income." The regression coefficient for the patient-level covariate is summarizing outcome differences of low income vs. non-low income patients at providers who are matched on the percent of low-income patients. The regression coefficient for the provider-level covariate "percent of patients of low income" is summarizing outcome differences for patients of the same income category who are treated by providers that differ with respect to their percent of low-income patients. If the patient-level covariate is negligible and the provider-level covariate is large, this is consistent with the interpretation that the association between income and outcome is related to systematic differences in qualify of providers who tend to treat more vs. fewer low-income patients (and not due to differences in outcomes for low versus non-low income patients within the same provider). In that case, care is needed because certain adjustment methods which fail to distinguish within- versus between-provider income differences may produce biased comparisons of providers. If the patient-level covariate is large and provider-level covariate is negligible, this suggests it may be important to adjust for income (to the extent that the mix of low-income patients varies across providers) and that failure to distinguish within- versus between-provider income differences may have negligible impact. If both patient-level and provider-level are large, then adjustment *methods should be used that remove the effects of within-unit differences (as they interact with varying unit proportions in the disadvantaged groups) but do not mask the quality differences among units.*

Examples in Literature

Feaster D, Brincks A, Robbins M, et al. Multilevel models to identify contextual effects on individual group member outcomes: a family example. *Fam Process,* 2011;50(2):167-183.
Abstract: This manuscript illustrates methods for utilizing measurements of individuals to identify group contextual effects on individual outcomes. Contextual effects can be identified by 1 of 3 methods: (1) divergence of the simple within- and between-group regression coefficients, (2) the presence of a cross-level interaction of the within- and between-group predictor variable, or (3) the effect of discrepancies within the group. These methods can be used to incorporate group context into an individual model and can be utilized for any individual process variable that might be affected by a group context. Example data include measures of hassles and coping adequacy of inner city, poor, African American new mothers, and their family members.

Reames BN, Birkmeyer NJ, Dimick JB, et al. Socioeconomic disparities in mortality after cancer surgery: failure to rescue. *JAMA Surg,* 2014;149(5):475-481.
Abstract: IMPORTANCE Disparities in operative mortality due to socioeconomic status (SES) have been consistently demonstrated, but the mechanisms underlying this disparity are not well understood. OBJECTIVE To determine whether variations in failure to rescue (FTR) contribute to socioeconomic disparities in mortality after major cancer surgery. DESIGN, SETTING, AND PARTICIPANTS We performed a retrospective cohort study using the Medicare Provider Analysis and Review File and the Medicare Denominator File. A summary measure of SES was created for each zip code using 2000 US Census data

linked to residence. Multivariable logistic regression was used to examine the influence of SES on rates of FTR, and fixed-effects hierarchical regression was used to evaluate the extent to which disparities could be attributed to differences among hospitals. A total of 596 222 patients undergoing esophagectomy, pancreatectomy, partial or total gastrectomy, colectomy, lung resection, and cystectomy for cancer from 2003 through 2007 were studied. MAIN OUTCOMES AND MEASURES Operative mortality, postoperative complications, and FTR (case fatality after >/=1 major complication). RESULTS Patients in the lowest quintile of SES had mildly increased rates of complications (25.6% in the lowest quintile vs 23.8% in the highest quintile, P = .003), a larger increase in mortality (10.2% vs 7.7%, P = .0009), and the greatest increase in rates of FTR (26.7% vs 23.2%, P = .007). Analysis of hospitals revealed a higher FTR rate for all patients (regardless of SES) at centers treating the largest proportion of patients with low SES. The adjusted odds ratios (95% CIs) of FTR according to SES ranged from 1.04 (0.95-1.14) for gastrectomy to 1.45 (1.21-1.73) for pancreatectomy. Additional adjustment for hospital effect nearly eliminated the disparity observed in FTR across levels of SES. CONCLUSIONS AND RELEVANCE Patients in the lowest quintile of SES have significantly increased rates of FTR. This finding appears to be in part a function of the hospital where patients with low SES are treated. Future efforts to improve socioeconomic disparities should concentrate on hospital processes and characteristics that contribute to successful rescue.

Appendix F: Illustration of Adjustment using Direct Standardization

With direct standardization, unit-specific rates are computed in each stratum and applied to a standard population case mix, producing an estimate of what might be expected if the provider were to treat the standard patient mix. You do not use a population average for the strata as in indirect standardization. This method sometimes becomes problematic if cell sizes are very small.

Table F-1. Illustration of Risk Adjustment using Direct Standardization

	All Patients in National Population		**Unit A**		**Unit B**		**Unit C**	
SDS Strata	Patient mix N/Percent	Clinically adjusted Deaths N/Percent	Patient mix n/Percent	Clinically adjusted Deaths n/Percent	Patient mix n/Percent	Clinically adjusted Deaths n/Percent	Patient mix n/Percent	Clinically adjusted Deaths n/Percent
	1,000,000 100%	22,000 **2.2%**	1000	22 **2.2%**	1000	26 **2.6%**	1000	29 **2.9%**
Average-High Income	800,000 80%	16,000 2.0%	800	16 2.0%	400	8 2.0%	400	8 2.0%
Low Income	200,000 20%	6,000 3.0%	200	6 3.0%	600	18 3.0%	600	21 3.5%
Provider rate for average-high income stratum applied to national proportion of average-high income				1.6%		1.6%		1.6%
Provider rate for low-income stratum applied to national proportion of low income				0.6%		0.6%		0.7%
Risk adjusted death rate accounting for sociodemographic risk - what might be expected if provider were to treat the national standard patient mix = SUM of provider stratum rate * national proportion for the stratum				**2.2%**		**2.2%**		**2.3%**

Appendix G: 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 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 5.
- 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 <u>not</u> 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 heart 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 separate document) 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 socicodemographic 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). 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. (Appendix C)
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 standard and reducing incentive 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 on p.14 – the text was correct, but the citation is Joynt, K. E., & Jha, A. K. (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 or plans 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 co-morbidities) or sociodemographic factors (e.g., recognizing and addressing non-English speaking). 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 could change the estimate of the provider's performance (either up or down). 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 methods paper.) 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) 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 a 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 the final report.

- ***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. The term "accurate" is also used sometimes to indicate precision and could be confusing. Language used to describe validity (to which risk adjustment relates) often varies by disciplines and preference. The references to correct or 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 the question on whether consideration of sociodemographic adjustment improves the performance measure for comparisons in performance and avoids incorrect inferences about quality. Although concerns about 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 P4P 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 of performance on the quality measure for the entity being measured. (i.e., without systematic deviation from the true value).
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 ask or demand 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. (See #7.)

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 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 use 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

a. 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.
b. Sociodemographic factors could be obtained from three sources:
    - socicodemographic 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.).
c. 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. Community factors could be used in at least two ways, each with specific conceptual and methodological considerations – 1) to assign a community characteristic to each individual patient (e.g., percent poverty or public transportation in community ***where the patient resides***) or to the healthcare unit (percent poverty or availability of public transportation in the community ***where the unit is located*** or ***for the patient population served***). The selected approach should meet the conceptual and empirical basis and sound methods for adjustment.

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.

a. 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.
b. 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.
c. Once finalized, the recommendations related to standing disparities committee can be inserted here.

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 this is outside NQF's role for endorsing performance measures and overlaps with the role of the Measures Application 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 at already underway at NQF to explore ways to make the measure endorsement and measure selection processes more coherent and efficient.